Authentic Assessment in the Classroom

Table of Contents

Chapter 1: Teaching, Motivation and Testing

School Contexts that Inspire Effort

Application

Authenticity
New Theories of Learning and Instruction
The New Cognitive Science of Learning
Metacognition
The Development of Good Thinking
Teaching, Motivation, and Testing
Authentic Learning Assessment
Summary

Activities
Suggested Readings

Chapter 2: Academic Motivation

Intrinsic Academic Motivation
Learned Helplessness in Children

Application

Attribution Theory
Self-efficacy Theory
Goal Theory
Authentic Learning Assessment and Motivation
  Task analysis
  Measuring what you task analyze
  Gathering and interpreting test scores
  Giving learners multiple-opportunities to improve
  Using test results
  Grading
Summary

Activities
Suggested Readings

Chapter 3: Developing a Framework for Authentic Learning Assessment

Establishing the Mission for an Authentic Learning Assessment Program
  Norm-referenced interpretations
  Criterion-referenced interpretations
  Growth-referenced interpretations

Application

Functions
  Setting expectations
  Diagnosis
  Monitoring learning
  Assigning grades

Application

Determining the Outcomes for an Authentic Learning Assessment Program
Bloom's Taxonomy
  Knowledge
  Comprehension
  Application
  Analysis
  Synthesis
  Evaluation
Gagne's Learning Hierarchies
  Verbal information
  Intellectual skills
  Cognitive strategies
Evaluation of Bloom's and Gagne's Systems
A Cognitive Learning Model for Determining Outcomes
  Declarative knowledge
  Procedural knowledge
Cognitive strategies
Metacognition
Motivational orientations

Activities
Suggested Readings

Chapter 4: Quality Assurance

Assuring Construct Validity
   Step 1: Specify the construct domain
   Step 2: Choose as many indicators of the domain as possible
   Step 3: Design the assessment task so that it requires only the
cognitive skills relevant to the construct you wish to
assess
   Step 4: Write task directions that require the thought processes
   that you intend to assess

Application

Gathering Evidence of Construct Validity
Instructional Validity
Assuring Instructional Validity
Gathering Evidence of Instructional Validity

Application

Consequence Validity
Reliability
Reasons Why Assessments Lack Reliability
   Limited sample of behavior
   Small number of observation occasions
   Unclear tasks
   Scoring imprecision
Assuring Reliability
   Increasing the number of performances
   Increasing the number of occasions
   Writing clear task specifications
   Increasing scoring objectivity
Summary: Quality Assurance Checklist
Activities
Suggested Readings

Chapter 5: Assessing the Knowledge Base

Knowledge Versus IQ
What is a Domain-specific Knowledge Base?
How Does a Knowledge Base Develop?
Assessing What Learners Know

Assessing Simple Factual Knowledge
Sampling
Recall Versus Recognition
Ways of Measuring Recall and Recognition
  True-false items
  Completion items
  Matching items
    Homogeneity
    Order of lists
    Easy guessing
    Poor directions
    Multiply correct responses

Application

Multiple-choice items
  Stem clue
  Grammatical clue
  Redundant words/unequal length
  All of the above/none of the above
Higher-level multiple choice questions
  Use justification to assess reasons behind an answer
  Use pictorial, graphical or tabular stimuli
  Use analogies to show relationships between terms
  Require application of principles or procedure

Application

Reliability
Ambiguous or unclear presentation of question
Not enough questions
Not enough time
Uncomfortable conditions
No test blueprint
Ambiguous answers
Construct Relevancy
Efficiency

Assessing Simple Understanding
The Restricted Response Essay
When Should Restricted Response Essay Questions be Used?
Scoring Restricted Response Essays
  Write good essay items
  Use Several restricted-response items
  Use a predetermined scoring scheme

Application

Sampling
Recall Versus Recognition
Reliability
Construct Relevancy
Efficiency

Assessing Knowledge Organization
Summary

Activities
Suggested Readings

Chapter 6: Assessing What Learners Can Do

Procedural Knowledge
How Procedural Knowledge Develops
  Cognitive stage
  Associative stage
  Autonomous stage
Types of Procedural Knowledge
Domain-specific basic skills
Domain-specific strategies
Domain-general strategies
Teaching Domain-specific Basic Skills
Assessing Domain-specific Basic Skills
Assessing Processes and Products
Designing the Assessment Context

Step 1: Clarify the procedure and the reasons for assessing it

**Application**

Step 2: Describe the assessment context
Specify the setting and level of intrusiveness
Specify who will observe the performance
Specify how much evidence is needed

**Application**

Step 3: List important behaviors and characteristics of the performance including common errors and mistakes

**Application**

Step 4: Design the procedure rating plan
Checklists
Rating scales
Holistic scoring
Anecdotal records

**Application**

Step 5: Arrange the behaviors or characteristics in an appropriate format

**Application**

Summary

*Activities*
Suggested Readings
Chapter 7: Assessing Problem Solving Strategies

Who Will Defend the Country: An Illustration
  Knowledge base
  Domain-general procedural knowledge
  Metacognition
  Domain-specific procedural knowledge

Assessing Domain-General Problem Solving Strategies
  Some General Principles for Assessing Problem Solving
    Novelty
    Authenticity
    Sustained, purposeful problem-solving
    Problems with multiple solutions
    Evidence of the problem-solving process
    Ratings

A Taxonomy of General Problem-solving Strategies

Teaching Problem Solving Strategies
  Up-front demonstration and explanation
  Mental modeling
  Guided practice
  Metacognitive information

Step 1: Identify and Define the General Problem Solving Strategy(ies)
  Cognitive strategies
    Analysis
    Comparison
    Inference and interpretation
    Evaluation
  Metacognitive strategies
    Planning
    Draft and tryout
    Monitor and revise
    Evaluate and reflect

Application

Step 2: Select the Context for Demonstrating One or More of the Above Strategies
  Cognitive strategies: Literature
Cognitive strategies: History
Cognitive strategies: Science
Metacognitive strategies: Science
Metacognitive strategies: Social Science

Application

Step 3: Design the Assessment Task
  Process versus product
  Visual maps
  Outlines
  Timelines, flowcharts, drawings, graphs, pie charts,
  genealogical charts
  Task instructions

Application

Step 4: Develop the Scoring Procedure

Application

Assessing Domain Specific Problem Solving Strategies
Identifying Domain-specific Problem Solving Strategies
  Physics
  Reading comprehension
  Writing
  Historical analysis
  Solving simple math problems: estimation
Step 1: Identify the Strategy
Step 2: Select the Context for Demonstrating the Strategy
Step 3: Design the Assessment Context
Step 4: Design the Scoring Procedure
Summary

Activities
Selected Readings

Chapter 8: Assessing Deep Understanding

The Essay Exam
Performance Assessment: Are These Activities or Tests?
Performance Tests Are Direct Measures of Learning
Performance Tests Assess Process and Products
Performance Tests Can Be Embedded in Lessons
Performance Tests Can Assess Social Skills
What Is Performance Assessment?

Developing a Performance Assessment
Performance Assessments Should Require Knowledge Construction
Performance Assessments Should Require Strategic Thinking
Performance Assessments Should Require Clear Communication
Performance Assessments Should Strive for Authenticity

Designing the Performance Assessment Task
Step 1: Decide on a Specific Subject Area
Application
Step 2: Define Those Cognitive Processes and Social Skills You Want to Assess
Application

Step 3: Design the Task and Task Context
Goal relevance
Level of difficulty
Multiple goals
Multiple solutions
Self-determined learning
Clear directions
Application

Step 4: Specify the Scoring Rubrics
Measure your goals
Select an appropriate scoring system
Checklists
Rating scales
Holistic scoring
Assigning point values

Scoring Rubric for Static Electric Performance Test
Checklist for social skills
Rating scale for cognitive processes
Rating scale for communication
Overall rating of the project

Application

Step 5: Identify Important Implementation Considerations
  Identifying testing constraints
    Time
    Reference material
    Other people
    Equipment
    Scoring criteria

Application

Delivering the assessment
  Structuring
  Motivational set
  Coaching
  Independent work
  Debriefing

Application

Summary

Activities
Selected Readings

Chapter 9: Assessing Genuine Achievement: The Portfolio

Act 1, Scene 1: Planning for Something New

Act 1, Scene 2: First Day of Class
  Rationale for the Portfolio
  Definition of a Portfolio
  Designing the Portfolio
Insuring Validity of the Portfolio
   Representativeness
   Rubrics
   Relevance
Step 1: Deciding on the Purposes for a Portfolio

Application

Step 2: Identifying Cognitive Skills and Dispositions

Application

Step 2: Deciding Who Will Plan the Portfolio

Application

Step 4: Deciding Which Products to Put in the Portfolio and How Many Samples of Each Product

Application

Act II, Scene 1: Third Week of Class
Step 5: Building the Portfolio Rubrics

Application

Step 6: Developing a Procedure to Aggregate All Portfolio Ratings

Application

Act III, Scene 1: Fifth Week of Class
Step 7: Determining the Logistics
   Timelines
   How products are turned in and returned
   Where final products are kept
   Who has access to the portfolio
A Final Conference

Activities
Suggested Readings
Chapter 10: Assessing Collaborative Skills and Self-Determination

Collaborative Skills
Teaching Learners to Collaborate
  Modeling
  Role playing
  Feedback
  Transfer
Taxonomy of Collaborative Skills
  Basic interaction skills
  Getting along skills
  Coaching skills
  Role-fulfilling skills

Assessing Collaborative Skills
Step 1: Clarify Your Reasons for Assessing Collaborative Skills

Application

Step 2: Describe the Assessment Context
  Specify the setting and level of obtrusiveness
  Specify who will observe and rate the collaborative skills
  Specify how much evidence is needed
Step 3: Identify and Describe each Collaborative Skills
Step 4: Design the Rating Plan
  Collaborative Skills Checklist
  Rating Scale for Collaborative Skills
  Holistic Rating of Collaborative Skills
  Narrative Summary of Collaborative Skills
Step 5: Develop a Scoring Form for Recording Your Ratings and
Arrange the Collaborative Skills in an Appropriate Format
Step 6: Plan for Giving Feedback to Your Learners

Self-determination
Defining Self-determination
  Setting goals
  Reaching goals
  Observing performance
  Rewarding oneself
Assessing Self-determination

Application

Activities
Suggested Readings

Chapter 11: Developing a Worthwhile Grading Plan

Why Assign Grades
Cultivating a Grading Philosophy
1. What meaning should each report card grade convey?
2. How should class grades be distributed?
3. What components go into a final grade?
4. How do you establish the importance of each component in the final grade?
5. How should you assign the grades?
6. Should students be given extra opportunities to raise their grades?

Application

Constructing a Grading Plan
Step 1: Check School District Policy
Step 2: Decide What You Want Each Grade Symbol to Mean
   Letter symbols
   Numerical symbols
   Limitations of the symbol systems
   Assigning meaning to the symbols
Step 3: Distinguish Reporting and Grading Factors
Step 4: Identify the Components of Your Grade
Step 5: Decide How Much Weight Each Component Will Have in the Final Grade
Step 6: Determine How the Components Will Be Combined into a Final Grade
Step 7: Choose a Method for Assigning Grades
   The fixed percentage method of assigning grades
   The total point method
Step 8: Decide What to Do about Borderline Scores
Grading Plan Checklist
Summary

*Activities*

Selected Readings
"Practice and you'll get better." We've all heard that advice. At times we may have believed that our practice would pay off, while at others, we were convinced there was nothing we could do to improve. Whether you are teaching world history, spelling, English grammar, French, or writing, you want your learners to believe that they are in control of their own learning. But experience convinces some learners otherwise.

Take a look into any classroom and you will notice some learners practicing as if their life depended on it, and others not even trying. To say that these students lack motivation would not be true. Observe them on the athletic field, in band, debate or drama where their desire to improve motivates them even when they receive no immediate rewards for success or consequences for failure.

What is it about some school tasks that motivate learners to high levels of effort, and other school tasks that cannot even capture a learner's attention. Could it be aptitude for the task? Maybe. But we've all seen learners who work hard to succeed at tasks they have little aptitude for, and others who fail to even try at tasks for which they have plenty of aptitude. To help explain the peaks and valleys of effort expended by learners, let's look at some of the contexts or settings in which a learning activity takes place.
School Contexts That Inspire Effort

The school settings where we see students expend high amounts of effort to learn and to improve are diverse: band, yearbook, athletic practice, drama rehearsals, debate, school newspaper, and academic activities relevant to their personal interests. Unmotivated learners are the exception rather than the rule in these contexts. But, visit some other settings and considerably more lethargy and passivity among learners can seen. What accounts for this difference?

School contexts that induce effort seem to have a number of things in common. Some of these are.

(1) **Students are assessed over what was taught and practiced.**

At the Saturday track meet, runners are tested over what they've been training for the previous week. At the Friday night band contest, performers are judged on how well they play the songs practiced during band class. Weekend debaters find that the knowledge and speaking skills acquired during weekly preparation determine the points they receive during competition.

But on the Friday history test, learners may see some questions that were not discussed or practiced in class. Biology students who spent the previous week dissecting animals, preparing lab reports, and making complex drawings may find themselves tested with true-false and multiple choice questions, which only remotely resemble the application tasks they've been assigned in lab. And, fifth grade math students may be puzzled about why the fraction questions on the test bear so little resemblance to those in the
textbook. What is so different about these two sets of examples? In the former students are certain of being assessed over what was taught and practiced, while in the latter they anticipate and may even expect their effort will be in vain.

(2) The focus of instruction is on solving problems and accomplishing tasks that are like those solved and accomplished by professionals in the field.

Students in journalism class do what real journalists do—they prepare a newspaper. Band students prepare for concerts—just as professional musicians do. Cosmetology students style hair; computer science students program computers. In these examples, the instruction is problem centered. In other words, what is learned in class serves some larger purpose. It is connected in some important way to their own prior experiences or to further learning in the same or more advanced class.

But in geography class, history class, math class, literature class, students may not always receive practice opportunities and complete activities that correspond with things they have already encountered in their world or that geographers, historians, mathematicians, or authors do in their daily lives. They may be learning facts and concepts whose mastery may appear to them to serve no larger purpose than to pass the chapter test. While geographers are involved in the conservation of resources, natural resource management, city planning, and map construction, in the classroom students are asked to memorize the 50 states, their capitals, major industries, and color this information onto dittoed map sheets.
While not all of the body of knowledge we call geography, history, literature, etc. has a direct relationship to a particular job or profession, the knowledge your students are asked to learn should bear a relationship to prior experiences and to some later learning—a relationship which should be pointed out and linked to applications beyond your classroom.

(3) **Standards or criteria of success are public, shared with the learner, and the focus of feedback.**

The sprinter on the track team knows what standards the coach and track officials use to judge her competence. Likewise, the cosmetology, journalism, or computer science student has little doubt about what is considered good performance. Thus, the standards or criteria of success that each learner is expected to live up to is openly shared. Suggestions for improvement or reasons for a poor performance are always related to the standards available for all to see.

This is not always the case in some science, social studies, or language arts classes. Learners in these classrooms may be puzzled at the grades and comments they receive on homework, class work, and test papers. Students may not know before a test what standards or criteria the teacher has in mind when scoring the test. And, these standards or criteria may be no clearer to the learner after the assignment is graded.

(4) **Assessment occurs over time during which the role of the teacher is to help the learner improve.**
You may have notice a great deal of conversation and consultation among learners and teachers during teaching and testing in school contexts characterized by high levels of effort. In fact, the lines between instruction and testing often become blurred in these contexts because assessment becomes a means of teaching and teaching becomes a means of assessment. Let's look at some examples.

In band class, is the Friday tryout for chair positions a test, or a training session? While a decision is made at the end of the tryout about who deserves the higher chairs, feedback is also given which allows the band member to improve her position next week. The same blurring between assessment and teaching may occur in a journalism class when the instructor gives the students repeated opportunities to improve his work. Is submitting copy or a first draft an assessment or is it an occasion for teaching? Are the Friday track practices that determine Saturday teams, the play rehearsals that determine roles, and the debate practices that determine partners instances of assessment or instruction? The answer is: they are both assessment and instruction.

Contrast the above with the end of chapter tests in history, geometry, literature, or chemistry. Tests are taken, points determined, grades assigned, and regardless of performance, usually the next chapter becomes the focus of instruction. Learners may or may not be given the opportunity to improve their knowledge, understanding, or skill after the test is returned. The line between assessment and instruction is clearly drawn.
(5) Learning contexts are provided where students show what they have learned under the same conditions that occur in real life.

In the world outside your classroom, writers are allowed to consult dictionaries, scientists to consult with colleagues, historians to have access to libraries, mathematicians to use calculators and other equipment to solve problems, researchers to have time to prepare, rethink, revise, and complete an experiment, student drivers are told in advance exactly what to expect during the road test.

During the typical classroom test, however, time is carefully controlled, no help is allowed from peers, learners are not allowed to consult dictionaries, textbooks, or their notes, and calculators and spell checkers may be prohibited. And, what is required may be known only on test day, and the scoring criteria clarified only when the test is returned.

(6) A hands-on exercise or problem is expected to be solved that produces an observable outcome or product.

In classrooms that encourage student effort the teacher assesses not only the product but the process used to arrive at it. Examples are the student newspaper, hair styling, solo trombone performance, lab experiment, tuned-up engine, or debate. These examples are bound together with the common thread that each results in an outcome or product that can be seen and examined according to established standards set by those who perform these activities in the world outside the classroom. They are "accomplishments"
that may include products (for example, poems, essays, charts, exhibits, etc.),
cognitive processes (for example, skills in acquiring, organizing and using
information), performances (for example, physical movements, as in oral
presentations, focusing a microscope, bisecting an angle, and gymnastics), or
attitudes and social skills (for example, habits of mind, group work and
recognition skills). These products and observable outcomes are testimony
that the effects of your teaching can be realized in a variety of ways.

APPLICATION

Think of an instructional context in which you were highly motivated to
learn. Now, review the six characteristics of school contexts identified above
that promote an effort to learn. Check below those characteristics from
among the six that were present in your context?

- Assessed over what was taught and practiced
- Assessed over tasks expected of professionals in the real world
- Knew the standards or criteria by which an assessment was being made.
- Assessed over time with time to improve
- Assessed under the same conditions that exist in the real world
- Assessed with a hands-on exercise or problem

List any additional characteristics not included in the six that motivated you
to learn?
Authenticity

One word that captures the six criteria above is *authenticity* (Wiggins, 1993). Authentic instruction and assessment identify the knowledge, thinking, problem solving skills, social skills, and attitudes exhibited by those in the community, on the job, or in advanced courses, as part of their normal work. Authenticity involves testing over what was taught and practiced in class and asks learners to use the same skills, knowledge, and thought processes modeled by adults at work, presented in class activities, covered in the text and workbook, and required outside the classroom.

Instruction and assessment are authentic when the tasks that are the focus of instruction and assessment are important tasks that tell the learners something about their knowledge and skills relative to themselves, rather than others. In other words, the learner's grades represent what he or she can do on some important task, not how far above or below average he or she is. Descriptions of performance are in terms of what the learner actually does.
relative to some standard of excellence rather than in terms of ranks, percentiles, or grade equivalent scores.

Authentic instruction and assessment, therefore, are designed to produce the learner's best rather than typical performance. Best performance occurs when learners are allowed multiple opportunities to demonstrate what they can do, and where each opportunity is followed by specific, performance enhancing feedback. Typical performance is that which occurs in assessment situations where only one opportunity to demonstrate learning is provided in which we must assume the learner gave his or her best effort.

To encourage a learner's best rather than typical performance during authentic instruction and assessment, the learner performs under a set of rules and regulations, called *constraints*, that match the conditions typically found outside the classroom. These include time to complete a task, reference material normally available to those performing the task, peer and adult support and consultation, equipment, prior knowledge of the task, and scoring criteria. These testing constraints are summarized in Table 1.1. We will have more to say about them in subsequent chapters.

Insert Table 1.1 about here

Finally, authentic instruction and assessment embody an outlook which says that the primary aim of testing is to inform and improve instructional decision making, rather than to label, grade, place or select the learner. Track meets are authentic because they inform the coach about what must be done in practice to improve the runners speed or endurance. Band contests are authentic because the results guide the band directors decisions.
about what to do during band practice. Likewise a biology instructor's tests become authentic when they have been designed to help make decisions about how best to conduct class so that learners improve their skill at observation or experimentation.

Authenticity is a concept that links teaching, testing, and motivation. The point of view of this book is that high levels of learner motivation will result when there is a link between what the real world expects of learners, what you expect of your learners, and what is asked of your learners. In the chapters ahead we will show you how to link teaching and testing in ways that will improve your teaching and create higher levels of motivation and effort in your classroom.

New Theories of Learning and Instruction

Motivation to learn may be linked as much to how we test as it is to how we teach. This approach to motivation is the result of research that has provided new insights on how children learn. In the remainder of this chapter we will discuss some of this research and the new directions for teaching and testing it has created. In the next chapter we will link this new perspective on teaching and testing to assessment practices that can motivate your learners to higher degrees of effort.

For most of the twentieth century theories of how best to teach and motivate children had little to do with how the mind works. This is because behavioral approaches to learning dominated most of this century's thinking about school instruction. The essence of these "behavioral" theories were that an understanding of how the mind works was not essential for bringing about
learning. Instead, learning was believed to occur from the following three conditions:

1. An environment scientifically designed to elicit correct and rapid performance
2. A focus on observable behavior or performance
3. Opportunities for feedback and positive reinforcement following behavior

These three elements make up the ABC model of learning, which is illustrated in Figure 1.1. The ABC model of learning refers to antecedents in the environment (A) that elicit desired behavior (B), which then becomes strengthened when followed by appropriate consequences (C). This model includes the essential elements of the behavioral science approach.

Insert Figure 1.1 about here

This approach to teaching and learning so dominated American educational psychology that not until the latter part of this century was it respectable for educational psychologists to theorize about and research the workings of the human mind. Now, after decades of research, a new approach to learning based on the processes of the human mind, called cognitive science, has achieved prominence and broad acceptance among psychologists and educators. Let's take a look at some of the learning principles behind this new approach.
The New Cognitive Science of Learning

New knowledge in the sciences of astronomy, geology, physics, chemistry, or medicine is often preceded by new technologies and new ways of conducting research. This also has been true in the science of learning. Since 1970, major advances have allowed us to study the way in which the brain actually works, that is, how experts in math, science, writing, and reading, etc. actually think, remember, problem solve, and learn, and how novice thinkers become expert thinkers (Brueer, 1993).

Cognitive approaches to learning are concerned with how everyday experiences are transformed or processed into mental images or sounds and stored for later use. In other words, they are concerned with how information is processed. It is logical, therefore, that cognitive psychologists have chosen the information processing model or computer to help describe the content and processes of good thinking.

Regardless of how they describe effective thinking and information processing, all cognitive approaches to learning share certain basic ideas. These shared ideas, or basic elements of the cognitive approach, shown in Figure 1.3, are:

- Relevant learner characteristics
- Instructional manipulations
- Cognitive processes
- Cognitive outcomes
- Outcome performance

While the behavioral approach to learning emphasizes how to establish a learning environment to produce more correct answers than incorrect ones,
the cognitive approach emphasizes that good thinking results from studying the cognitive processes and outcomes that underlie right and wrong answers.

Insert Figure 1.2 about here

Much of what we know about the content of thinking and assessment today is owed to advances in cognitive science. Let's look at some of these advances in three areas: Knowledge, cognitive strategies, and metacognition. Then, we will point out the significance of this understanding for classroom testing and motivation.

Knowledge

Studies of the differences between how experts in reading, writing, math, social studies, or science think, remember, and learn in comparison to novices highlight the importance of knowledge. Thinking, whether in American History, second grade reading comprehension, geometry, fourth grade writing, or life science springs from a well organized and easily accessible knowledge base.

This knowledge base takes two forms: domain-specific knowledge, and procedural knowledge. A domain-specific knowledge base contains the facts, concepts, rules, and generalizations pertaining to a specific area or topic. Learners who can think about the Civil War, perform first grade addition and subtraction, complete a writing assignment, play chess, or predict gravitational forces know a lot about these domains. Learners who
know little can't think as effectively when faced with problems to solve in these areas regardless of their motivation or aptitude (Ceci & Liker, 1986).

Good thinkers in any area also possess procedural knowledge. They know how to quickly and automatically perform the actions required for effective writing, typing, weighing, focusing a microscope, setting up lab equipment, booting a floppy disc, or doing triple column subtraction with regrouping.

Novice thinkers, on the other hand, lack the domain specific knowledge and procedural knowledge necessary to skillfully problem solve. When a problem is presented to them, they lack the facts, concepts, or other information necessary to properly understand or make sense of the task. Making sense of the task involves the learner constructing a mental idea or representation of the problem, as shown in Figure 1.3. It is this mental representation of the task or problem that is prerequisite for effective thinking and problem solving.

Insert Figure 1.3 about here
(Revise art of student constructing a mental idea using number line example in text below.)

The notion of mental representation is crucial for understanding the cognitive science of learning. Representations are the ideas, images, or words you construct in your head to think about what you see or hear. When you ask first graders to add the numbers four and two, successful learners use the mental number line to help them compare and count. They hear or read
the words "What is four plus two?", visualize a number line in their heads which allows them to see that four is greater than two and then they "count-up." In other words they say to themselves or speak out loud, "four (pause), five, six."

This mental number line is part of the domain-specific knowledge that expert adders possess, while the counting strategy is part of their procedural or know-how knowledge. Cognitive scientists have learned that unsuccessful adders don't yet possess the ability to represent the addition problem in terms of the mental number line. Only when they acquire this knowledge, do they become successful at simple addition.

Thus, learners who have little prior knowledge (whether domain-specific or procedural) when first learning will have difficulty learning anything new. Middle school children who lack knowledge about friction and gravity will have a difficult time correctly representing force and motion problems in physics. High school students will understand little of what they read in the chapter on photosynthesis if they have little knowledge about cells and cell structure.

Cognitive psychologists believe that learners, even at the earliest grade levels, have some knowledge about nearly every topic they study. This information may be in the form of ideas, however vague, unconnected facts, implicit rules, or images. And often it may be wrong--what cognitive psychologists call misconceptions. These misconceptions can be found in domain-specific knowledge where a first grade learner believes that small objects like sand or sugar have no weight. Most high school juniors and seniors have the faulty rule that when two objects of different weight are dropped or thrown horizontally, the heavier one always hits the ground first.
Such prior misconceptions will affect how they represent problems of force and motion.

Prior procedural knowledge can also be faulty. In the 1970's Brown and his colleagues found that many children approach multi-digit subtraction with a set of procedures which render them virtually helpless (Brown and Burton, 1978). They identified 80 simple bugs in their procedures like switching to addition if the top number is smaller than the bottom number in a subtraction column. When faced with a zero to borrow from, they change the zero to a 9 but forget to do anything else. The problem looks right but gives the wrong answer.

Sometimes misconceptions are so entrenched in the learners' minds that they continue to use mistaken ways of thinking and behaving even when alternative methods have been taught (Roth, 1990, 1991). And, sometimes prior beliefs are so strong that learners ignore statements that they disagree with or they choose not to believe what they see.

The implications of this research for teaching and testing are clear:

(1) Before teaching something new, know what important facts, concepts, principles, generalizations, and procedures are needed.

(2) Assess the extent of both correct and incorrect prior knowledge.

(3) Teach correct ways for learners to represent what is taught. (4) Assess both correct learning and the processes used to get there.

Extensive, well-organized, and easily accessible domain or procedural knowledge is a necessary but not sufficient condition for good thinking and learning. We know this from studying how expert writers, readers, chess
players, or historians go about doing their business. They plan, model, draw 
analogies, search memories, underline, take notes, ask themselves questions, 
elaborate on what they hear or read, and monitor their understanding as they 
are listening or reading, practicing, and rehearsing.

Cognitive learning specialists use a variety of expressions to refer to 
these types of thinking skills: learning-to-learn skills, general thinking skills, 
reasoning skills, problem solving skills, or cognitive strategies. They all 
refer to general methods of thinking that improve learning across a variety of 
subject areas. When learners use cognitive strategies they are engaged in 
mental activities that go beyond the processes that are naturally required for 
carrying out a task.

For example, as you read this paragraph you are engaged in decoding 
processes (eye movements, sounding out phonemes, searching memory for 
meaning, and so on). Decoding is not a cognitive strategy because it is 
naturally required for reading. However, if before you began to read this 
chapter you scanned the headings and asked yourself questions about the 
chapter's content, and if as you were reading you regularly paused and asked 
yourself if you understood what you were reading, then you would be using 
cognitive strategies to help your reading comprehension.

Similarly a student is not using a cognitive strategy when he regroups 
and borrows to solve a subtraction problem—he is doing what is naturally 
required to perform the task. However, if before solving the problem and 
during the act of problem solving, the learner prompted himself with 
statements such as, "What am I supposed to do? What information am I 
given? First, I'll draw a picture of what the problem is asking," he would be 
using a cognitive strategy.
There are cognitive strategies to improve memory, reading comprehension, writing, math problem solving, and spelling. Throughout most of the 1970's and early 1980's, cognitive learning specialists believed that these general skills and reasoning abilities, like asking questions, planning, and memory rehearsal, would improve a learner's performance across all areas of the school curriculum. Since these strategies could be used in any area of specialty, they are referred to as domain-independent learning skills, as contrasted with domain-specific knowledge and procedural skills. These domain-independent skills were believed to be a principle component of intelligence (Sternberg, 1989). Once learned they would transfer to successful learning in any academic field.

However, research carried out in the 1980's convinced cognitive researchers that extensive experience and knowledge in a specific domain like reading, chess, or biology, was more important for learning than general thinking or cognitive strategies (Pressley, 1995; Gardner, 1993). The most recent research (Bruer, 1993; Sternberg, 1996) suggests that schools should teach both extensive domain specific knowledge and general thinking skills, and that the development of subject matter expertise requires both types of instruction. This new synthesis draws one other important conclusion: learners must learn to control and monitor their own thinking and learning—a process call metacognition.

Metacognition

Knowing how to use a cognitive strategy to improve learning in math, writing, or history is no guarantee that a learner will use it (Pressley, 1995). Learner's must also be aware of when to use the strategy and to monitor how
it is working. When students think about how they think, are consciously aware of themselves as thinkers, are aware of the thinking strategies that they use, and ask themselves how these strategies are helping them, they are engaged in mental processes called metacognition. Teachers usually control the learning of cognitive strategies, but learners must control their use. Metacognition is the conscious control by learners of the use of cognitive strategies.

This knowledge, awareness, and control of thinking processes develops with age. Researchers have shown that children in early elementary school can engage in metacognition (Pressley, Borkowski, and O'Sullivan, 1984, 1985). However, their ability to do so is not learned automatically. For example, although using one method for improving spelling works better for a group of learners than another, the learners aren't aware that one strategy is better than another unless this is explicitly pointed out.

Along with extensive exposure to domain-specific and procedural knowledge, and the teaching of cognitive strategies, the new science of cognitive learning tells us that learners must be taught metacognition. Thus, if you want your learners to use cognitive strategies to improve their reading, spelling, math computation and problem solving, etc., you must teach them three things: domain specific knowledge and procedures, cognitive strategies, and metacognition. This latter process involves teaching learners to: (1) attend to the effectiveness of strategies, (2) attribute differences to the relative effectiveness of a particular strategy, and (3) make a commitment to use the more effective strategy in future decision making. Otherwise, learners may not use a given strategy, not notice whether it is effective, or fail to use it when they should.
The Development of Good Thinking

Expert thinkers in any area use general thinking or cognitive strategies to solve important problems. They plan, practice, model, draw analogies, evaluate, and rehearse. They are also consciously in control of these processes. They are aware of themselves as thinkers, monitor how their thinking is going, and evaluate how they are improving or learning. They apply these cognitive strategies and control their use on a foundation of an extensive, well-organized, and deeply understood domain-specific knowledge base. General strategies need a knowledge base from which to operate.

Your learners will be novice thinkers who are in the process of becoming expert thinkers—or at least more competent ones. They will pass through three overlapping but distinguishable phases of learning during this journey to mastery (Anderson, 1983). As we discussed above, the initial stages of learning in any subject area are characterized by vague ideas, misconceptions, and rudimentary procedural know-how skills. As you teach them information, they gradually transform their prior knowledge into new concepts and generalizations. During the second phase of learning, their domain specific knowledge accrues and becomes better organized. Their know-how or procedural knowledge allows them to perform more fluently. They gradually get better at controlling the use of cognitive strategies. Finally, the third phase of learning is characterized by deep understanding in the knowledge domain, the automatic and efficient use of procedures, and the flexible use and control of cognitive strategies.
Classrooms that produce high levels of learner effort monitor the development of these building blocks of good thinking as conscientiously and systematically as a swimming coach monitors the development of lung capacity, swimming stroke, kicking and turning techniques. The challenge for assessment is the development of instruments and processes that allow teachers to track the growth of learners' thinking as they move from the initial states of domain-specific learning, to the knowledge compilation, restructuring, and fine tuning phase, to the final stages of expertise, competence, and mastery. In subsequent chapters we will show how to accomplish this in your classroom.

Teaching, Motivation, and Testing

We have reviewed what the new science of cognitive learning says is important for the development of good thinking in any subject area. Now we must ask "What are the implications of this new perspective on learning for teaching, motivation, and testing?" Let's begin by examining the work of two researchers whose program exemplifies this perspective on learning.

Griffin and Case (1975) set out to create a program to develop successful math learners. Griffin and Case were particularly interested in readiness for first grade math instruction. Such instruction usually begins with teaching learners rudimentary addition and subtraction skills. They concluded that the knowledge about numbers that learners bring to first grade is a principal determinant of whether they master first grade math skills or experience problems learning math throughout the early elementary school years. Their research focused on how to identify or assess correct or incorrect prior number knowledge (prior to learning how to add), identifying
the new knowledge and thinking strategies required for the successful learning of addition, developing ways to teach new knowledge and thinking skills, and then assessing the extent of learning.

Griffin and Case began their work with American children living in economically deprived inner-city communities. They administered tests of conceptual knowledge of math and found that these children entered kindergarten and first grade significantly behind their middle-class peers in the conceptual understanding and knowledge of facts and procedures that are minimally necessary for success in early addition and subtraction. Moreover, they found that much of their prior-knowledge was "non-adaptive"—that is, it prevented the learning and development of strategies that would make them successful at basic computation. They developed the Rightstart program to remedy this.

The Rightstart program began with a comprehensive analysis of the knowledge and skills required for correctly solving addition and subtraction problems. Successful kindergarten and first grade math students can count, know which of two numbers is larger or smaller, which numbers come before or after others, and which numbers are closer or farther away from a given number. But, most importantly, they think about or represent simple addition or subtraction problems in terms of a mental number line. Children who conceptualize quantities on a number line can think about numbers in a continuous fashion. Those without such a representation see numbers in a global or polar fashion. Their quantitative thinking is restricted to ideas like big things are worth more than small things but don't perceive a continuous stream of numbers.
From this task analysis, Griffin and Case developed a test to assess the number knowledge and strategies of novice math learners. They next developed a curriculum to teach learners to think in terms of a mental number line, and to use adaptive strategies to solve simple computation problems. They created a series of thirty interactive games which involved hands-on activities for learners to acquire necessary number knowledge.

Learners were regularly tested on their development of this knowledge. At the end of the program, low SES learners who experienced Rightstart far surpassed control groups. But more significantly, at the end of first grade the math skill of these low SES learners equaled that of their middle class peers.

The Rightstart program illustrates a seamless link between teaching and testing. Testing informed teaching: it identified the knowledge and skills learners needed to be successful at computation. Teaching flowed from testing: it focused on developing the knowledge and skills which tests showed that learners lacked. Testing was conducted to help develop a learner's skill and not to simply label a learner as deficient and assign a grade. It provided diagnostic information to teachers to help them better focus their instruction.

Moreover, children learned trying to produce results. Each test produced concrete evidence of their learning and development. Each lesson was linked with the test. In most cases the lesson was the test--the teacher taught, coached, provided practiced, and observed and recorded improvement. The procedures involved in such assessment helped convince learners that they were in control of their learning--that if they try they would improve.
Rightstart is just one of many educational programs whose instructional and assessment practices derive from cognitive learning theory. For example, another program, Thinker Tools, is a physics curriculum used in middle schools to change some of the misrepresentations or naive ways of thinking about ideas such as forces, motion, and gravity. It was developed by Barbara White, a cognitive psychologist, and Paul Horwitz, a physicist (White and Horwitz, 1988). It combines innovative assessment and instructional techniques that have been shown to motivate learners and help them acquire scientific thinking skills.

Instruction and assessment of reading comprehension has been advanced significantly by the work of cognitive scientists such as Campione and Brown (Campione and Brown, 1990). Likewise grade school writing (Scardamalia and Bereiter, 1986), middle school social studies (McKeown and Beck, 1990), and high school geometry (Wertheimer, 1990) programs illustrate the new cognitive science of learning and its effects on teaching, motivation, and testing.

These effects have been to provide new purposes for classroom teaching and testing. Cognitive science has identified the reasons behind why some learners think and learn better than others. Rather than attribute such differences to intelligence, cognitive science tells us that good thinkers in any area possess deep domain-specific knowledge, cognitive strategies and metacognitive skills. Differences in these areas underlie learners' performance differences. Effective school instruction focuses on these areas as goals for student learning.

Thus, cognitive science proposes a new theoretical basis for classroom testing: classroom tests should assess learners' performance changes in what
they know and how they think about specific subjects. These results should allow teachers to track the growth of learners' knowledge bases, and thought processes. They should provide diagnostic information to teachers about what needs to be learned and how it should be taught. Cognitive scientists have a new term for such testing that distinguishes it from traditional measurement practices: they call it learning assessment (Glaser, Lesgold, and Lajoie, 1987).

Authentic Learning Assessment

We have been using the terms "measurement," "testing," "evaluation," and "assessment" interchangeably. But, now we will be more precise to show how learning assessment is different from what typically goes on in classrooms and schools.

When we measure, we typically assign values or numbers to represent something we cannot see: knowledge, understanding, ability, skill, attitude, or learning. We can measure these attributes in a variety of ways: observation, rating scales, homework, class work, oral questions. But most commonly we do so with paper and pencil tests. These tests can be short quizzes, long chapter tests, or final exams. Thus, testing is a vehicle by which we measure attributes we can't see.

When we add up the number right on such tests we are still engaged in measurement: Sally got ninety percent correct, Larry got 19 of 23 spelling words correct, Alicia earned 50 points on her homework. But during measurement we don't associate any value with this number such as good/bad, intelligent/unintelligent, desirable/undesirable, acceptable/unacceptable. As soon as we assign some type of value to that
number, which usually happens during grading, we are engaged in *evaluation*. During evaluation we assign a grade of "A" to the 90 on a history test and this becomes associated with goodness, or desirability, or intelligence. Conversely when we assign a "C" grade to an essay that earned 7 out of 10 points, we associate that 70 percent with "average," "mediocre," or "OK."

Thus, classroom *measurement* involves determining a quantity to describe some attribute of a learner or group of learners, *testing* is the vehicle for doing so, and *evaluation* is the value or importance we place on that measurement. Then what is assessment?

*Assessment* is what cognitive psychologists refer to as an "ill-defined concept." A well defined concept is like a table: you can point to one and you know what it is. It has easily recognizable attributes: legs, a top, a particular shape. Ill-defined concepts are much harder to define. Love is a good example--as is assessment.

The best way to think of assessment is as a process that involves a lot of things done at different times, rather than as one thing done at one time (like giving a quiz). The most important steps in this process are:

1. Identifying the important knowledge and thinking skills necessary for good thinking in a particular subject area.
   This is what Griffin and Case did as part of *Rightstart*.
2. Developing a variety of ways to measure the knowledge and thinking skills identified in (1). These can involve observation of actual behaviors, oral questioning, paper-and-pencil tests, writing assignments, attitude scales, etc. Such instruments affect daily and weekly instructional planning and adaptation.
(3) Collecting information from all these sources.

(4) Interpreting this information. To what extent is the learner acquiring the knowledge and thinking skills that are the focus of instruction? How automatic is the learner's use of acquired knowledge and thinking skills? We don't assign a number or grade at this point. Rather we develop "thick" descriptions of performance that serve to diagnose learning problems and inform subsequent teaching.

(5) Using this information to guide further teaching: Based on the teachers' interpretation of the learner's performance on the tasks developed in (2), the teacher alters teaching practices, provides more practice, more feedback, or gradually fades involvement, giving the learner more independence.

(6) Repeating steps 3-5 until the learner has acquired the knowledge and skills necessary for competence.

(7) Assigning some type of summative grade or description of learning that traces the development of the student's learning from the beginning or novice stages to the competence stage.

As you can see, learning assessment carries with it more meaning than measurement, evaluation, or testing. It involves planning, developing, measuring, interpretation, diagnosis, feedback, and evaluation as part of the instructional process.

Summary
We began this chapter by posing a problem: How do we convince learners that, if they only put out the effort, they will learn? We proposed that motivation to learn comes from a connection between teaching and testing. We used the term *authenticity* to characterize this unbroken bond. We then reviewed the theoretical underpinnings of the importance of such a bond when we described the new science of cognitive learning and its principal findings. We concluded by pointing out that this new science of learning not only provides a theoretical and research base for new approaches to teaching, but also for new approaches to classroom testing. We used the term *authentic learning assessment* to refer to these new approaches. In the next chapter we will show the crucial link between motivation and learning assessment. After reading and studying that chapter, you will be able to begin authentic learning assessment in your classroom.

Activities

1. Describe a particular learning context that you have experienced that encouraged you to produce your best rather than typical performance. How did it differ from the many other instructional contexts you have experienced?

2. Contrast the behavioral "ABC Model" of learning with the newer cognitive science model of learning. Can you think of an instructional context in which each would be appropriate?

3. In your subject or grade provide one example each of domain-specific knowledge, procedural knowledge and metacognitive knowledge.
Drawing on your own experience, how would you describe the differences among these three types of knowledge.

4. Recall Griffin and Case's RightStart program to develop successful math learners. What principles of cognitive science did their program include that helped make it a success?

5. In your own words, describe the difference between the terms "measurement," "testing," and "evaluation." What do these three concepts have in common with "assessment?"

Suggested Reading


A highly readable and substantive presentation of the contributions of cognitive science to our understanding of classroom learning.
Table 1.1. Some Of The Most Common Testing Constraints

**Time.** How much time should a learner have to prepare, rethink, revise and finish a test.

**Reference material.** Should learners be able to consult dictionaries, textbooks, or notes as they take a test?

**Other people.** May learners ask for help from peers, teachers, and experts as they take a test or complete a project?

**Equipment.** May learners use computers or calculators to help them solve problems?

**Prior knowledge of the task.** How much information about the test situation should learners receive in advance?

**Scoring criteria.** Should learners know the standards by which the teacher will score the assessment?
Figure 1.1. ABC Model of Learning
Figure 1.2. Basic Elements of the Cognitive Approach to Learning
Chapter 2
Academic Motivation

In the previous chapter we pointed out that a serious challenge facing teachers is convincing students that effort will produce learning. How to do this is one of the most puzzling problems facing educators. One difficulty in understanding how to motivate learners is our uncertainty about the meaning of motivation. Is motivation an inherited trait, something we are born with and over which we have little control? Or, is motivation a term we use to describe behavior that can be easily influenced by rewards and consequences?

In this chapter we will discuss intrinsic motivation for academic tasks. We will examine recent research on what promotes and underlies this type of motivation. Consistent with the perspective on learning and assessment presented in Chapter 1, we will view academic motivation from a cognitive framework. Our belief is that the learner's thought processes, more than rewards, influence classroom motivation. These thought processes are strongly influenced by classroom experiences, the most important of which are testing practices. In this chapter we will show how authentic learning assessment is a key to invoking the effort your students will need for learning.

Intrinsic Academic Motivation

Intrinsic motivation is what gives energy and direction to behavior in the absence of rewards and punishments. In classrooms, intrinsic motivation influences learners to choose a task, get excited about it, and persist until they master it, regardless of whether they get a reward for doing so. The
first grader who practices handwriting because she likes to see neat, legible letters like those displayed on the letter chart is intrinsically motivated. The fourth grader who puts together puzzles of states and countries because he likes to see the finished product and wants to learn the names of capital cities is intrinsically motivated. So, too, is the ninth grader who repeats typing drills because he likes the feel of his fingers hopping across the keys, and connects that sense with the sight of correctly spelled words on the page.

Surveys of children's beliefs and expectations about learning as they first enter school show high levels of optimism and enthusiasm (Pressley, 1995; Eccles & Wigfield, 1992). Children come to school expecting that they will be successful, get good grades, and be liked by their teachers and peers. These same studies also show a significant decline in motivation for school subjects starting in the middle elementary school grades and continuing until high school (Anderman and Maehr, 1994). The most serious declines in interest and enthusiasm for academic activities seems to occur in the sixth and seventh grades. At this time significant numbers of boys and girls decide that they can't succeed in school and that effort will do them no good. Once acquired, these beliefs resist change and undermine later academic achievement (Fincham, Hokoda, & Sanders, 1989).

Cognitive learning theorists, particularly in the 1990's, have expended much effort accounting for this decline. From their effort a consensus has emerged as to the role of thoughts, beliefs, and self-perceptions in motivating learners. This perspective replaces earlier views which held that motivation was dependent on needs and drive states influenced primarily by rewards and punishment.

This new motivational paradigm holds that the effort that students put into studying and mastering academic subjects is a result of their self-
perceptions and expectations for success primarily determined by experiences in the classroom. Something happens between the early primary grades and middle school that convinces many learners that nothing they do will earn them better grades or the esteem from teachers and parents that they seek.

Theorists like Jacquelynne Eccles speculate about a developmental mismatch that occurs around the sixth grade between a learner's psychological needs and the demands of school (Eccles et al., 1993). As learners approach the middle school grades, they have increased needs to be liked, accepted, make decisions, become autonomous, experience success, and feel as though they have control and power over their lives (Hartup, 1989). Schools, especially middle schools on the other hand, allow learners fewer opportunities to make decisions, sometimes group learners by ability or practice other forms of tracking, emphasize standardized tests, allow fewer opportunities for learners to improve on tests, and grade more rigidly than primary schools. The net effect of these school conventions is a decrease in intrinsic motivation leading to passive school behavior.

The origins of this new model of motivation can be traced to a series of experiments conducted at the University of Pennsylvania by Martin Seligman and his colleagues (1969, 1967). Let’s look at what they found.

Learned Helplessness

Seligman and his counterparts trained dogs in an escape/avoidance task using a rectangular box that had a shoulder-high barrier in its center. The floor of the box was made of heavy gauge chain-link fencing through which a mild electric current could be passed. A dog was placed in the cage and a 10-second warning light (in the dogs line of vision) was occasionally
presented, followed by a 50-second electric shock delivered through the floor. If the dog jumped over the barrier while the shock was on, the shock immediately ceased; if it jumped before current was sent to the floor, the light was turned off and the scheduled shock canceled. Thus, the dog could either escape a mild shock by quickly jumping across the barrier when it first felt the current, or it could avoid the shock altogether by jumping when the light came on.

Seligman trained several groups of dogs using this procedure (Seligman and Maier, 1967; Seligman, 1969). One group was experimentally naive—they were placed in the box with no prior experience of its features. But, a second group was given preliminary training in the box. They were restrained in a harness and given electric shocks which they could not escape. No matter what movement the dogs made, they could not escape the shock.

The naive dogs, when first placed in the box, did what any dog would do when the shock was felt—they ran about until they eventually scrambled over the barrier and escaped the shock. After only several experiences or trials of this kind, the dogs escaped more quickly and eventually received little or no shock.

But the behavior of the second group of dogs who received preliminary training of inescapable shock was very different. When first placed in the box and upon feeling the shock they behaved just as the naive dogs did—they ran about and whined. But, after about 30 seconds they typically would lay down on the floor until the 50 seconds of shock was up. They simply gave up. And for most of this group, this passive acceptance of discomfort continued on subsequent trials. They showed no signs of learning to escape or avoid the shock or motivation to do so.
Even more puzzling was the behavior of those dogs in this second group who did manage to occasionally jump over the barrier and escape the shock. When a naive dog did this, a single success was enough to permanently learn jumping behavior and the dog would quickly repeat this response on all succeeding trials. But the dogs given inescapable shock during the pre-training experience showed no sign of learning from a successful escape. On the trial following a successful escape, they immediately reverted to the pattern of passive acceptance that they previously displayed.

Seligman (1969) hypothesized that the second group of dogs formed a belief or expectation during the pre-training experience of inescapable shock. He called it a "hopelessness set." They learned a belief that nothing they did would terminate the shock. On one occasion, purely through happenstance, they might lift their hind leg just before the 50-second shock period was up. From that they might form the expectation that leg lifting terminated the shock. But they would quickly learn on the next trial that leg lifting or any other behavior would not have the effect of ending their discomfort.

Seligman proposed a "cognitive" explanation for the dogs' behavior. They formed a belief or expectation which caused them to behave passively when shocked. Seligman's explanation for the results of his first experiment was called the "learned helplessness" hypothesis.

But, there is another explanation for the behavior of the second group of dogs--a non-cognitive one. Their lack of motivation could have been due to their repeated experience of intense shock. This experience resulted in a conditioning of fear to the box so that when the dogs were placed in it, they were immobilized by fear. With this explanation, exposure to intense shock
and not the inescapability of shock (which caused a hopelessness set) could explain lack of learning and motivation.

To find out, Seligman and Maier (1967) gave a third group pre-training and shock that was inescapable—just like the second group but with one difference. While confined in the harness, they could terminate the shock by pressing a panel with their nose or head. Each dog in this third group was trained with a dog from the inescapable shock group—a procedure called yoked pairs. Whenever a dog in the third group pressed the panel and terminated the shock, the shock was also terminated for his buddy in the second group. Thus, both groups of dogs received the same amount and intensity of electric shock.

Remember that the alternative explanation for the behavior of the dogs in the second group that received inescapable shock was that exposure to shock alone made them behave helplessly. Were this explanation correct, then you would predict that after pre-training, both the second and third group of dogs would display passive behavior when they were shocked in the box. Both groups received equal amounts and intensity of shock.

But the cognitive explanation would predict just the opposite. According to the cognitive explanation, the third group of dogs learned that shock was escapable—that something could be done to terminate it. The second group, however, learned no such expectation. They still believed that shock was uncontrollable. Which hypothesis was correct? As Seligman predicted, the third group of dogs learned to escape shock as quickly as the first group of dogs, while the second group again displayed learned helplessness. Thus, dogs were not simply learning a behavior as a result of their experiences—they were learning an expectation.
Cognitive motivational theorists speculate that certain school-wide and classroom practices, in particular, testing and grading practices, create learned helplessness in many children. Just as Seligman's pre-trained dogs learned the expectation that nothing they do can terminate shock, some learners acquire the belief that nothing they do will result in better grades. The experiences that develop this hopelessness are those described in the previous chapter: (a) poor grades on tests that may not bear a direct relationship to what is learned and practiced in class and on homework and (b) feedback in the form of a single overall test score without opportunity to correct mistakes or master what was not learned. After a series of such tests, poor grades without an opportunity to improve convince some learners that there is little they can do, and so they stop trying.

Standardized tests, particularly high-stakes exams like the Iowa Test of Basic Skills, National Merit Qualifying Exam, Scholastic Assessment Test present a similar situation to some learners. Every year they take these or similar tests which may not bear a direct relation to what they learned and practiced in class. Poor scores on these exams have important implications for classroom grouping, admission to special programs, admission to college, and other important decisions in the life of the learner. Learners who consistently score below average on these tests may conclude that nothing they do can improve their scores.

APPLICATION
Identify a personal testing or assessment situation in which you were convinced that nothing could be done to make things better. Then, list below the circumstances that caused you to adopt this disposition of helplessness. Looking back, what were some conditions that if changed might have helped avoid your feeling of helplessness.

Circumstances promoting a feeling of helplessness:

________________________________________

________________________________________

________________________________________

________________________________________

Conditions that might have been changed:

________________________________________

________________________________________

________________________________________

________________________________________

Learned Helplessness in Children

Dweck and Reppucci (1973) wondered whether, as a result of certain school experiences, some children believe that there is nothing they can do to improve their performance. To test their hypothesis, they had an adult experimenter give a group of fifth graders easy math problems while another adult experimenter gave the same group unsolvable ones. The children had 2-8
a series of math sessions with both adults but in no particular order or sequence. After the initial training (notice the similarity with Seligman's experimental procedure), the adult who gave children unsolvable problems (the failure experimenter) gave the children some simple problems that they had already solved for the adult who gave them easy problems (the success experimenter).

The surprising result (but not surprising for those familiar with Seligman's work) was that the children could not solve the easy problems given to them by the failure experimenter and, just like Seligman's subjects, many didn't even try. The interpretation of their findings was that the children had learned to be helpless from the failure experimenter.

The early work of Seligman and Dweck and Reppucci laid the intellectual foundations for the new cognitive paradigm of academic motivation. They demonstrated that environmental stimuli or events, particular contexts, or our experiences don't exert a direct impact on our efforts to learn. Rather, they affect our cognitions, self-perceptions, beliefs, or expectations for success which influence motivation. This shift to a cognitive model of motivation has resulted in three distinct yet overlapping theories to understanding academic motivation: attribution theory, self-efficacy theory, and goal theory. Let's examine the ideas contained in each and their implications for teaching, motivation, and testing.

Attribution Theory

At this point in your life you've succeeded at some things and failed at others. Think about one of your more recent successes or failures. Were
you successful or unsuccessful because of the amount of effort that you exerted, your natural ability, luck, or some combination of these? Another way to ask this question is: Do you attribute your success to internal forces, such as effort or ability, or to external forces, such as luck or the difficulty of the task?

Weiner (1986), a leading proponent of attribution theory, believes that people naturally seek to understand why they succeed or fail. Students, for example, when asked to explain why they received a certain grade, typically refer to their hard work or effort, innate ability, an easy or hard test, or luck. These causes originate either within the learner (effort, ability) or outside the learner (luck, task). Locus of causality is the term applied to the origin of these causes: internal locus of causality or external. This three-dimensional model of attributions is shown in Figure 2.1.

Insert Figure 2.1 here

Motivational theorists like Weiner assert that only when learners attribute their success to effort are they likely to exert genuine effort to study hard for a test. If learners attribute their success or failure to ability, luck, or task difficulty—all of which are out of their control—they are like Seligman's dogs believing that nothing they can do will improve their grades. Thus, how a student thinks about or interprets success or failure, and not the experience of the outcome itself, determines the energy and direction of their efforts.

There is evidence, especially for students with learning disabilities, that attributional beliefs affect academic motivation. Learning disabled
students are more likely to attribute poor achievement to low-ability than normal achieving peers (Jacobsen, Lowery & DuCette, 1986). Similarly, low achievers who assert that effort is the primary factor in their success, achieve more than their low achieving peers who believe that ability is the reason for lack of learning (Kistner, Osborne, & LeVerrier, 1988).

The importance of attributions for academic motivation raises the question of what teachers can do to convince students that they are in control of their learning. Much of the research in this area addresses things that teachers do to convince them otherwise. For example, grouping learners by ability sends them the message that this uncontrollable trait is important to the teacher and should be important to them (Slavin, 1987). Assigning grades based on comparisons with other students—a practice called competitive or norm referenced grading—appears to have a similar affect on learner attributions (Spaulding, 1992; Gagne, Yekovich, and Yekovich, 1993.)

Dweck (1975) studied the effect of attribution retraining (i.e. persuasion) on effort attributions. She identified a group of children who made ability attributions and gave them success experiences in math problem solving followed by a few encounters with failure. After each failure experience, the experimenter said things to the children that attributed the failure to low effort. These children were better able to deal with future failure than a comparable group of learners who only experienced success. Apparently the failure experience and persuasive comments served to inoculate the learners against loss of future motivation.

But, the most successful attribution retraining programs recognize that counseling students to attribute their success to effort would do little for low achievers if their efforts didn't result in improved learning. Consequently,
these programs combined attribution retraining with training in the use of
cognitive and metacognitive strategies and assessed learning with tests that
measured both the outcomes of learning and the mental processes used to
achieve those outcomes.

Borkowski and his colleagues (Borkowski, Weyhing, & Carr, 1988)
have worked with low-achievers in reading who make ability attributions to
explain their failure. They taught these students reading comprehension
strategies, while at the same time emphasizing that they could understand
what they read better if they would use metacognitive strategies showing
learners that their reading comprehension scores were improving as a result
of their use of these strategies. Their program exemplified both the new
cognitive model of academic motivation and the close relationship between
teaching and testing relationship discussed in Chapter 1.

Two other groups of students participated in Borkowski's study. One
group learned strategies but did not receive attribution retraining, another
group received neither treatment. The group that received the full program
when tested three-weeks following training, demonstrated far superior
reading comprehension and use of cognitive strategies than the other two
groups.

Thus, research on academic motivation using attribution theory
presents convincing evidence that attributions are important causes of effort
and that attributions can be retrained. But, the key to doing this is teaching
in a way that convinces learners that school success is largely under their
control. This involves instruction in cognitive strategies, attribution
retraining, and assessment practices linked to instruction that show learners
their progress in learning both products (domain-specific knowledge) and
processes (cognitive strategies).

2-12
Self-Efficacy Theory

Self-efficacy theory holds that academic motivation hinges on learners' beliefs that they can succeed at school tasks. Bandura, the originator of self-efficacy theory, defines self-efficacy as, "people's judgments of their capabilities to organize and execute courses of actions required to attain designated types of performance" (Bandura, 1986, p.391). In other words, students are more likely to begin, persist at, and master tasks that they think they are good at. This judgment is what is meant by self-efficacy.

Judgments about self-efficacy differ from attributions. Attributions are perceived causes of success or failure. They influence expectations and behavior. They are one type of information that learners use when making a judgment about self-efficacy. If a learner believes that success in calculus is due to being born with math ability, and believes that he possesses little of such ability, he will have low self-efficacy for calculus.

Bandura identifies several other sources of information that learners use to make judgments about self-efficacy. One is verbal persuasion where the teacher expresses faith and confidence to learners that they can be successful. Another is seeing peers succeed at a particular task. If a learner sees someone whom they like and admire receive high marks or praise from a teacher for solving a difficult geometry theorem, she is more likely to believe that she can do likewise.

Perhaps the most important piece of information used by learners when making self-efficacy judgments is past experiences of success or failure with a particular task. The learner who has received high marks for his three previous essays, will have higher self-efficacy for the next writing.
project than the learner who consistently earns low grades. Thus, a student weighs a variety of information in addition to attributions when coming to a judgment of self-efficacy for a particular subject. Once made, the judgment is a principal cause of the level of effort and persistence that a learner expends on learning as well as the level of achievement obtained.

Self-efficacy is not a stable personality trait, however. There is no self-efficacious student. Rather, judgments of self-efficacy are subject specific. The same student can have high self-efficacy for reading, but low self-efficacy for math.

Teachers can play an important role in a learner's self-efficacy for a specific subject area or task. This is because they control the difficulty level of the task, the nature of instruction to teach the task, and the assessment techniques that provide learners with information to judge their success or failure.

Task difficulty is an important consideration for teachers concerned with increasing self-efficacy. Students master easy tasks quickly. They make rapid progress towards competence on activities that are below their skill level. And, this provides them with little information about their abilities and skills. Conversely, learning goals and tasks that are beyond the learner's skill or ability level produce failure and diminish self-efficacy. Ideally, teachers should challenge learners with tasks and goals that are just beyond their skill level. How do teachers select tasks and goals that have just the right degree of difficulty? A task analysis of lesson content, which we will describe shortly, and assessment of learner skills prior to instruction based on task analysis are required.

Have attempts to enhance self-efficacy for particular tasks been successful? There are several lines of research that seem to suggest this.
Jacquelynne Eccles (1985, 1989) has documented that girls have lower self-efficacy for math than boys. Her observations of classrooms where girls had favorable math self-efficacy in comparison to classrooms where they did not underscores what we have said above. Girls have higher self-efficacy in classrooms that do not use competitive grading practices, where ability comparisons are minimized and where there is continuous assessment of performance based on hands-on activities.

Other researchers studying self-efficacy (Schunk, 1991; Bandura, 1988; Elliot and Dweck, 1988) document increased self-efficacy when teachers help learners set challenging yet realistic goals with their students. Likewise self-efficacy is enhanced when teachers set specific performance indicators along with these goals (e.g. type 30 words per minute rather than increase your speed; or spell 90 percent of words correctly). Continuous performance feedback based on performance assessments sustain and heighten self-efficacy. This latter finding underscores the importance of decreased reliance on quizzes and tests as the only means of evaluating learners.

Goal Theory

Goal Theory is the third and most recent perspective that derives from the cognitive motivational paradigm. Let's return to some of the early work of Carol Dweck (1975, 1973) to gain a fuller appreciation of this approach. Recall that her initial work with fifth-grade math learners showed that uncontrollable failure created in the learner's mind a disposition of helplessness which caused them to make no effort to solve easy problems to which they had already given correct answers.
In the late 1970's Diener and Dweck began investigations about whether the goals for learning set by children had a more or less pronounced effect on the development of learned helplessness (Diener and Dweck, 1978, 1980). They administered psychological tests to groups of children to determine their academic goal orientation. They found that children generally fall into two goal groups with regard to interest in learning: one group--called the task-focused group--focused on developing competence in the academic area and improving their skills for purely intrinsic reasons. The other group--called the ability-focused group--engaged in learning tasks with the goal of "showing-off their ability, outperforming others, and gaining external rewards like praise and good grades."

In the Diener and Dweck studies they asked both goal groups to solve 12 math problems. The first eight were easy for all children to do but the remaining four were clearly beyond every child's skill level. The researchers asked the learners to think out loud as they were solving the problems expressing both their thoughts and feelings. They also asked the children to make predictions about the likelihood of successful problem solving both before and after they first experienced failure.

Both groups of students showed little differences in what they thought or felt about the tasks and future success on the first eight problems. But, when failure was encountered, dramatic differences between the task-focused and ability-focused groups emerged. Ability-focused students tried to avoid finishing the problems, guessed randomly, and became bored and anxious. They doubted their success when only a few seconds before on the easy problems they were quite optimistic. In other words, they mistrusted their ability.
The task-focused children weren't even aware that they were failing. When they thought out loud, they saw the difficult problems as opportunities to learn, tried to use cognitive strategies and metacognition to solve them, and expressed optimism about their own ability and likelihood of solving the problems. Many of them actually began using more sophisticated strategies as they persisted in problem solving. They liked the tasks and enjoyed the challenge.

When asked to explain and interpret the episode, the two groups made very different attributions. The ability group attributed failure to lack of intelligence while the task group voiced a need to work harder and think better. While the former group saw nothing they could do to get better and preferred to avoid such tasks in the future, the latter group wanted more opportunity to improve and work harder.

This study and others by Dweck (1988) and her colleagues show that the two categories of goals are distinct and result in dramatically different degrees of academic motivation. Table 2.1 displays a range of meaning for the different goals.

---

Insert Table 2.1 about here

---

Much of the research on goal theory has shown that students who adopt task-focused or mastery goals are more likely to achieve in school, make more use of cognitive strategies when problem solving, expend a lot of
mental effort searching their memories and relating new learning to prior learning than ability oriented learners.

This research also shows that classroom instructional practices have a strong influence on the goal orientation that learners adopt. One of the principal researchers in this area is Carol Ames (Ames, 1990a). Her research documents that as children proceed through the elementary school grades, they increasingly believe that their ability sets a limit on what they can learn. Certain classroom activities exacerbate this viewpoint. Ability grouping is one of them. Such a practice is correlated with learners' developing an ability rather than a task focus to learning.

Similarly, students are more likely to adopt an ability orientation in classrooms that emphasize getting the highest grades or being the best. Such classrooms post grades, display the best projects, nominate only certain students for awards, etc. Some students in these classrooms may become more interested in getting the work done and getting a grade than in learning.

Ames also documents that certain testing practices contribute to an ability focus. Standardized testing that results in norm-referenced scores by which we compare children with one another reflect and enhance an ability orientation. An ability focus is promoted by classroom testing and grading practices characterized by student comparisons, percentile and percentage scores, student rewards for grades and little feedback on how to improve one's test performance.

Thus, goal theory places special emphasis on classroom practices as a determinant of a student's personal goal beliefs. These goal beliefs enhance judgments of high or low self-efficacy. These judgments, as we have seen,
also affect a broad range of student motivational behaviors like persistence, use of learning strategies, choices, and preferences.

Ames work also shows that the goal beliefs of students can change as a result of teaching practices. Ames worked with a large number of elementary teachers of at-risk learners to change their goal orientation. She assembled a notebook of strategies to enhance task-focused goals. These strategies included suggestions for how learners were grouped, instructional strategies, the nature of feedback, test construction and interpretation, and grading practices (Ames, 1990b). Results from a full year of using task-focused learning and assessment strategies show that at-risk learners perceived their classrooms as being more task-focused, had more positive attitudes towards school, had higher academic self-concepts, showed an enhanced preference for challenging work, reported using more effective cognitive strategies, and were more intrinsically motivated than control students.

Learning Assessment and Motivation

The cognitive-motivational paradigm represented by attribution, self-efficacy and goal theory focuses on beliefs as the principal causes of academic motivation and effort. Attribution theory relates learner effort to their explanations for success and failure. Self-efficacy theory stresses learner's judgments of success or failure as reasons for working hard in specific subject areas. Goal theory tells us that learners who take a task-focused perspective on school learning exert more effort and engage in more successful learning behaviors than those who take an ability orientation.

The cognitive paradigm of motivation also identifies school and classroom instructional practices that affect a learner's effort, strengthening
judgments and beliefs. Our interest is in those classroom evaluation practices that can augment or detract from effort attributions, judgments of self-efficacy, or task-focused goals.

In Chapter 1 we outlined a process called authentic learning assessment that promotes these three motivational cognitions. Let's examine each component of the authentic learning assessment model and see how it does so.

**Task analysis.** Our learning assessment model begins with an activity called task analysis (Gagne, 1985). In general, you carry out task analyses when you identify the goals, objectives, and activities for your lessons and analyze the important knowledge and thinking skills required by learners to be successful. Students are more likely to be successful with lessons that have been task analyzed than with those that have not. This is because task analysis informs you of the precise skills and information required for task mastery and this information provides clearer direction for your teaching. It also makes clearer to learners the steps involved in mastery and allows them better to monitor their progress. Figure 2.2 illustrates a task analysis for a lesson on expository writing.

---

Insert Figure 2.2 about here

---

Thus, task analysis makes it more likely that learners will experience success quickly in a subject area and, as we have seen, this is a principal determinant of self-efficacy. In addition, task analysis helps you identify the learning strategies which promote success and high-self-efficacy. With 2-20
respect to goal theory, task analyses engender in learners the notion that competence in reading comprehension, writing, solving equations, or figuring out physics problems is a matter of mastering the skills required of the task rather than having a particular aptitude or ability for math, science, or writing.

**Measuring what you task analyze.** Your task analysis will identify a variety of learnings required for student mastery. Some will be cognitive: learning facts, understanding rules, acquiring important concepts and generalizations, using learning strategies. Others may be non-cognitive: developing certain attitudes, habits of mind or social skills. While some of these types of learning will be best measured by paper-and-pencil tests, others are better measured by observation, performance tests, projects, checklists, rating scales, or interviews.

Students feel less anxious about learning and have increased self-efficacy and task-focus in classrooms that find ways to evaluate some learning without relying on traditional quizzes and paper-and-pencil tests (Hill, 1984). Deci and associates (Deci, et al. 1991) strongly advocate the use of projects to evaluate learning. Their research documents enhanced intrinsic motivation (which they call self-determination) in classrooms where learners focus their efforts on accomplishing complex projects and where they receive continuous feedback as they work to complete it.

Most likely you will use a variety of sensory modalities to determine what your learners know: verbal, non-verbal, writing, oral explanations, drawings, bodily movements and gestures, etc. Ideally, the measures you use to assess varieties of learning should use the same sensory modalities, and involve the same behaviors as the activities you used to teach them. That

2-21
way learners will make the connection between what they do in class and how they are to perform on a test. They will perceive the testing situation as something they have been prepared for which will lead to positive self-efficacy judgments. Moreover, techniques that assess learning in different ways (for example, quizzes, take-home problems, and projects) will allow learners who are stronger in some abilities than others to show what they have learned in ways that are most compatible with their strongest modalities, thus enhancing their expectations for success. They will also connect the amount of effort expended in learning with the degree of effort required to perform. This should reinforce attributions about the importance of effort in their success.

Moreover, they will see that the test or assessment technique is an indicator of how much they learned and how much they tried during class. When a test is perceived to be unrelated to the time and energy expended during class and on homework, they may be confused as to its purpose. More often than not, they will perceive the only purpose of assessment is to grade and rank—a perception that leads to an ability rather than to a task goal orientation (Anderman and Maehr, 1994).

**Gathering and interpreting test scores.** As Ames (1990b) has shown, testing practices that focus on grades, rankings, or comparisons with other learners (percentiles) promote an ability goal orientation. Conversely, interpreting scores in terms of how much learning has occurred and how much learning remains puts the focus on learning, progress and skill development.

Learners pass through three overlapping stages in their journey to competence: an initial stage characterized by vague prior knowledge and
misconceptions, a transition stage in which their knowledge base becomes increasingly structured and skills more efficient, and a mastery stage typified by deep understanding and the automatic use of procedures and strategies (Snow, 1989). Interpreting test scores to learners in terms of their progress through stages promotes effort attributions, self-efficacy, and task focus.

**Giving learners multiple-opportunities to improve.** As we pointed out in Chapter 1, school settings that motivate students to high levels of effort allow them many opportunities to improve their performance. Classroom assessment practices in the academic areas should do likewise. This allows them to observe their own progress over the course of lessons. When the practice of multiple assessment opportunities is combined with teaching learners how to monitor and record their own progress, self-efficacy and task-focus are increased.

**Using test results.** Learners want to get better at what they do. Just observe them playing video games and you will be convinced. Test outcomes that simply result in a grade frustrate this innate desire for competence. Grades serve an evaluation function--they make a final value judgment about the extent of learning. But, they don't typically tell learners why they have performed well or performed poorly. That's the function of feedback. Feedback provides them with information about the quality of their performance and what they need to do to get better. Feedback includes specific suggestions for improvement.

Some classrooms confuse the function of grades and the function of feedback, believing that the former is a substitute for the latter. While the function of grades is to classify, label, or make a summative judgment about
a learner, the purpose of feedback is to show learners how they can become more competent. Assessment practices that produce continuous feedback, followed by multiple opportunities to apply it in repeated assessments encourage improvement, progress, success, and exert positive effects on attributions, self-efficacy, and task-focus (Spaulding, 1992).

Test results should be interpreted not only for what they say about student learning but also for what they say about effective teaching practices. They should help diagnose both student learning problems and problems in delivering instruction. Ideally, they should lead to altered instruction, more effective instruction, learner success, and enhanced academic motivation.

**Grading.** Cognitive-motivational theories do not suggest that grades are irrelevant or harmful. Grades will always play a significant role in the assessment process. Students not only want to progress and to become competent, but also to have someone say that they did so. This is also true for their parents. Grades are the principal means by which schools inform them of their children's accomplishments. But, often summative grades like a "B+," "needs improvement," or "93%," may not do justice to what a learner has accomplished or otherwise failed to do.

Grades can also distort a learner's accomplishments. Usually grades represent an average of the points earned on tests, quizzes, homework, chapter tests, projects and final exams. Because they are an average, grades may conceal the degree of competence displayed on the individual components of the grade on which the average is based. Although Steve and Stacy both earned an 87 for their final social studies grade, an 87 doesn't mean the same thing for these two learners. Stacy's 87 may represent
excellent accomplishments in conceptual understanding, average learning of facts, and below average progress in the writing of her ideas. Steven's 87 might reflect a different pattern of accomplishment. Average grades encourage assigning the same degree of accomplishment to learners with the same average, whether it is a number, letter, or percentage.

Thus, from a learning assessment perspective, final grades should be supplemented with what are called "thick descriptions" of performance (Gipp, 1995). Such descriptions not only identify the category or value assigned to a student, but point out the different types of accomplishments that are the reasons for the category or grade.

Summary

Motivation is a topic typically left out of books dealing with teacher testing practices. However, testing and motivation are inextricably linked. In the remainder of this book we will describe an approach to classroom assessment that is different from what typically occurs in many schools. As we will see in the next chapter, an understanding of an approach called authentic learning assessment and the decisive role of motivation on learner effort and performance will change the way you teach as well as how you test.

Activities

1. Provide a specific example from your experience in which you were "intrinsicly motivated." In your opinion what role did self-perceptions and expectations for success play in your motivation to learn?
2. Following the research of Seligman, describe a specific situation in which someone's response to in a learning situation could be described as "learned helplessness."

3. Using what you know about attribution theory, self-efficacy theory, and goal theory, describe how each theory could account for why a student does poorly in math. Identify a learner characteristic that distinguishes each theory from the others.

4. This chapter introduced some of the components for implementing an "authentic learning assessment" in your classroom. Using a subject-matter example, describe what you would have to do in your classroom to implement this model of assessment.

Suggested Reading


The article reviews some of the most recent and most thoroughly researched perspectives on academic motivation. It emphasizes how these research findings are applied to adolescents.
Chapter 3
Developing a Framework for Authentic Learning Assessment

After several weeks in your first teaching assignment, you will notice some of your learners flourishing, some making steady progress, and others struggling just to keep up. Inevitably, you wonder why. Just as some students attribute their achievement entirely to ability, so do some teachers. You may attribute differences among your students to intelligence, aptitude, innate talent, or gift. But these explanations will prove inadequate for two reasons: (1) Many students with similar abilities and aptitudes show vastly different degrees of learning, and (2) these explanations fail to provide teachers with practical alternatives for helping learners of all ability levels.

In Chapters 1 and 2 we identified reasons for why some students thrive in school and others don't. We pointed out that motivation and the link between teaching and testing are essential ingredients of student learning. But, of equal importance for explaining differences in achievement are the cognitive mechanisms underlying it. Cognitive learning theory identifies four vehicles for learning: knowledge, cognitive strategies, metacognition, and motivation. Consequently, the goal of your teaching should be to assess growth in these areas over the school year. The purpose of this chapter is to help you organize a blueprint for doing this.

A Theory To Guide Your Authentic Learning Assessment Program

Since you are going to assess student learning, you will need an understanding of what it is you are assessing. In this chapter we will introduce some important themes in cognitive learning theory to help you.
These themes will give purpose to your assessments and help integrate the information you collect into a coherent picture of learner achievement.

Cognitive learning theory is a family of learning theories that seeks to answer three fundamental questions: What is the content of good thinking? What happens to this content as learners progress from their initial exposure in an area of learning to competence? And, what type of teaching best promotes this development? Learning assessment focuses on the first two questions.

Figure 3.1 shows the important ideas raised by these first two questions. Looking down the first column of Figure 3.1, we see the five elements that make up good thinking discussed in Chapters 1 and 2. These were declarative knowledge (or, domain specific knowledge), procedural (or, know-how knowledge), cognitive strategies, metacognition, and motivational orientation (e.g. attributions, self-efficacy, goal focus). Notice that, unlike ability, aptitude, or intelligence, these elements can be influenced by your teaching.

Insert Figure 3.1 about here

Running across Figure 3.1 are phases or periods on the continuum from novice to expert. Down the columns are states of learning and transitions that assessments such as observation, paper-and-pencil tests, interviews, performance tests, student think-alouds, and cognitive maps are designed to measure.

Figure 3.1 suggests that assessments of learning should use a variety of assessment methods to trace a learner's progress in five areas of thinking.
from initial naive states to desired states of competence and expertise. So, if your principal were to ask the question, "What student behaviors are you assessing in your classroom?" and you refer to Figure 3.1 for help, you might answer:

*I am assessing my learner's progress from partial knowledge to deep understanding, from hesitant and uncertain application to the automatic, efficient, flexible and adaptive use of procedures, their use of cognitive strategies and metacognition, and their motivation, beginning with their willingness to try and ending with their valuing of effort.*

While you may be promising your principal a lot, this chapter will give you the understanding to fulfill your promise. But, before doing so, let's look at several traditional ways of conceptualizing classroom assessment and compare them with the learning assessment you have promised.

Establishing The Mission For Your Authentic Learning Assessment Program

What is the reason for developing an authentic learning assessment? What is the purpose for spending significant amounts of time developing a variety of assessment tools, administering them, and then collecting, organizing and communicating the information they provide? Although the result of these activities will usually be a report card grade, your purpose for assessment should go far beyond grading. Formulating such a mission statement involves a variety of considerations. Some of these involve an articulation of your unique educational philosophy, an understanding of community and school values, and your own thoughts on how learning
occurs. To help you formulate a coherent purpose for your assessment, we will introduce two important concepts: frame-of-reference and function.

Frames of Reference

Your assessment techniques will generate a lot of information about students. How do you want this information to be interpreted? Or, in other words, what will be your frame of reference for making sense out of the numbers, scores, percentages, and descriptions of performance that your tests or observations produce? Teachers typically employ three frames of reference when making sense out of assessment information: norm-referenced, criterion-referenced, and growth-referenced.

Norm-referenced interpretations. When social psychologists use the term norm, they refer to a group standard of conduct that governs how the members of the group behave. There are group norms that tell people what they should wear, where they should sit at a wedding, how they should cut their hair or speak and act.

When the word "norm" is used in assessment, it refers to the average group performance on a test. If the class average on a spelling test were 83, and you received a 92, a norm-referenced interpretation would be, "You scored above average." With this interpretation, you know little about how much you know or how expert you are in the subject over which you have been tested. But, you do know that you did better than average.

Norm-referenced interpretations of learner performance occur routinely when standardized tests are used. A student who scores at the 56th percentile is someone who did better than 56% of those who took the test. We don't know the degree of mastery of math or reading achieved by a
learner who scores at the 56th percentile. But, we do know that it is slightly above average.

Teachers often make norm-referenced interpretations of classroom performance. They do this when they compute each students score for a test, rank them from highest to lowest, and assign a grade based on where the student ranks in the class. When a teacher assigns a grade of "C" to the middle 50% of the group, she is giving a norm referenced interpretation of student performance.

Students who receive norm-referenced grades know where they stand in the class, but typically not how competent they are in solving algebra equations, or writing essays. One result of norm-referenced grading practices is that they tend to foster a competitive orientation to school learning. They may also mask degrees of mastery or improvement of specific domains of knowledge, skills and competencies, since they represent an average performance over many specific skills.

**Criterion-referenced interpretations.** Criterion-referenced interpretations are like those given to members of the track team: "you ran the mile in 5 minutes and thirty two seconds." Or, like those given to a journalism student: "your first draft contained two grammatical errors," or a typing student: "you typed 45 words a minute with two errors." These interpretations require that a teacher specify various criteria of mastery for a particular area. After students take the "test," their performance is described in terms of how well they met the criteria specified—not how far above or below average they are.

Criterion referenced grading occurs when various categories or grades (A, B+, etc.) are associated with different degrees of student mastery. For
example, a grade of "A" means that a student can type better than 60 words a minute with no more than one error; a "B" is given to learners who type between 50-60 words a minute with no more than two errors, etc.

As you can see, criterion-referenced interpretations of test scores tell learners what they can do. They have the potential for fostering the effort attributions and task-focused goals discussed in Chapter 2. The problem for the teacher is determining the degree of mastery required. This is often determined arbitrarily. No such problem exists for norm-referenced interpretations: the benchmark of success is relative to the average performance which is provided by the performance of those who took the test. In other words, a standard external to the test is not required for assigning scores and making interpretations.

**Growth-referenced interpretations.** Most teachers would like to acknowledge not only learning but also growth or improvement in learning. They would like the learner who started third grade knowing no multiplication facts to be rewarded for having mastered most of them on the six-week exam more than a learner who already knew the multiplication tables at the start of the third grade.

But, growth-referenced interpretations are typically conveyed informally to learners and their parents in conversations or comments made on report cards. They rarely are presented as a grade. Consequently, improvement in learning, although highly valued by parents and teachers, rarely becomes a basis for assigning grades.

Over the years, there have been attempts by measurement experts to develop techniques for including growth in assigning final grades (Kubiszyn & Borich, 1996). Although these techniques require additional assessment,
some authors believe growth-referenced interpretations should be used as supplements to norm or criterion referenced interpretations of learning.

APPLICATION

In the space below, write a brief comment about what you would say to parents at the first open-school night of the year about the "frames of reference" they should use to interpret the information you send home about their child's performance on your classroom assignments, tests, and homework.

Classroom assignments:

____________________________________________________________________

____________________________________________________________________

____________________________________________________________________

____________________________________________________________________

Classroom tests:

____________________________________________________________________

____________________________________________________________________

____________________________________________________________________

____________________________________________________________________

Standardized tests:

3-7
Functions

Whether you adopt a norm-referenced, criterion-referenced, or growth referenced frame of reference, or a combination of these, your choice will be influenced by the function or purpose your data are to serve. Grading is one of those functions. Traditionally four functions have been assigned to classroom assessment: setting expectations, diagnosis, monitoring learning, and assigning grades.

**Setting expectations.** Most teachers seek information about learner performance before instruction begins. Traditional classroom testing programs often use this information to set expectations for individual students. For example, knowing that a learner's math skills rank in the bottom 25% of your class at the start of the school year, will influence the expectations that you have for what that learner will achieve. In doing so,
you are making *norm referenced* interpretations of your learners' performance.

*Criterion-referenced* interpretations can also be used to establish expectations. For example, let's say that a teacher's final goal for a three week third grade multiplication unit is that a learner will solve multi-column multiplication problems with 95% mastery. This teacher will have different expectations of mastery for those students who know their times tables and math facts before the lesson in comparison to those learners who don't.

Like criterion-referenced assessment *growth-referenced* assessment makes use of learner performance prior to instruction, but, not for setting expectations. The basic assumption underlying growth-referenced assessment is that all learners will learn. The purpose of growth-referenced assessment is to help them do so—not to set expectations for how much or how little growth will occur. Consequently, growth-referenced assessment uses pre-instruction information from tests or observations to discover factors that may hinder progress.

**Diagnosis.** Diagnosis can occur prior to, during, or after instruction to determine the source or potential source of a learning problem. In many schools and classrooms, diagnosis invariably begins after a learning problem has been detected. It will usually be carried out by someone other than the classroom teacher, such as an educational diagnostician or school psychologist. In these cases, diagnosis focuses on determining weaknesses in the abilities or aptitudes that are presumed to underlie successful reading, spelling, writing, math problem solving, and so on. Norm-referenced interpretations are made of the data generated by these diagnostic tests.
Some specialists also use diagnostic tests that assess whether the learner has the prerequisite skills to spell, write, or read effectively. In such cases the assumption is that the learner is experiencing learning problems because he or she isn't prepared for the instruction. Criterion-referenced interpretations are made of the results of these tests.

Growth-referenced assessment also uses measurement for diagnostic purposes but with two important differences. In growth-referenced assessment, diagnosis is carried out by the teacher to uncover learning problems prior to and during instruction—not after it. As you will recall from Chapter 2, cognitive learning theory places emphasis on what learners know in a specific area prior to instruction. The purpose of instruction, therefore, is to develop, organize, and increase the complexity and efficiency of the learner's knowledge base. Diagnosis helps the teacher determine what misinformation or misconceptions, or "buggy" math rules the learner brings with him or her to instruction. Instruction and assessment, then, focus on correcting the misinformation, debugging and amending the math rules, and modifying the misconceptions. Growth-referenced assessment assumes that learning problems are caused by misconceptions or inadequacies in how learners think about or represent problems, not in terms of abilities or prerequisite skills.

**Monitoring learning.** In traditional classroom testing programs, teachers monitor learning by making informal observations and asking questions. The focus of monitoring is on the products of learning or observable outcomes, such as precisely spelled words, answers to math problems, grammatically correct sentences, accurately read sight words, or correctly answered reading comprehension questions. Ideally, such
information should be used to decide if instruction in an area should cease, be altered in some way, or continue. But, teachers may monitor learning more for reasons of giving praise and feedback to students than for what it may indicate about their instruction. Both criterion- and growth-referenced interpretations are made of the data gathered from monitoring.

Monitoring learning is an integral part of growth-referenced assessment. But, in contrast to more traditional testing practices, growth-referenced assessment uses monitoring, and focuses on the processes underlying learning as well as on the observable performances or products which suggest that learning has occurred. Growth-referenced assessment begins with an analysis of the knowledge and thinking skills required for competence in a particular subject area. How the student becomes competent is as important as the proof of competence.

Consequently, authentic learning assessment uses specific questioning strategies, verbal prompts and hints, and techniques called student "think-alouds" and "cognitive-maps" during instruction to systematically monitor the learning and use of cognitive strategies and the increasing complexity and organization of the knowledge base. The information gained from these techniques provides feedback to learners that enhances motivation, but also feedback to teachers to revise, continue, or start a new lesson.

Assigning grades. Nearly all classrooms assign some number, letter, or label to represent the sum total of a students learning at the end of a term or semester. These summative judgments can derive from a criterion-referenced or norm-referenced (also called "grading on the curve") interpretation of assessment data. We identified a number of problems with such grading practices: they assume equal amounts of learning have
occurred for individuals who achieve the same grade, fail to acknowledge continuous progress or development in learning, and may mask an individual student's learning strengths and needs.

Growth-referenced assessment can result in a letter grade—if that's what is required by a particular school or school system. But growth-referenced assessment requires that a learner's performance be described in terms of what was learned, how it was learned, and the changes that came about in the learner's knowledge base and use of cognitive strategies. Thus, "thick descriptions" of learning take precedence over single indicators of achievement whether in terms of averages, percentiles, letters, or category labels.

Now, having reviewed the frames of reference and functions of assessment, let's develop an overall mission statement which expresses the outcomes you desire from your assessment of student learning.

APPLICATION

In the space below, write a brief statement of what you would say to parents at the first open-school night of the year indicating the relative emphasis you will place on each of the following functions of assessment.

Setting expectations:

________________________________________________________________________

________________________________________________________________________

3-12
Determining the Outcomes For an Authentic Learning Assessment Program

The focus of your assessments will depend on the theory of learning that underlies your teaching. Unfortunately, most textbooks that cover how to identify goals and objectives for teaching don't always make this clear. The result is that some teachers teach and construct an assessment that doesn't match the kinds of learning they value.

As we have seen, traditional approaches to assessment derive from behavioral learning theory (Sulzer-Azaroff, et al., 1988). This theory holds
that the focus of instruction should be on observable behaviors and skills. These learning skills are hierarchically sequenced such that more advanced skills cannot be learned before less advanced ones. There is a correct sequence for teaching these skills that applies to all learners. Learning involves the accumulation of these skills. The difference between more complex and less complex learning and problem solving is a matter of the number of observable skills that the learner has mastered. According to behavioral learning theory, teaching and assessment should focus exclusively on these observable outcomes and not on the unobservable cognitive processes (e.g. the use of cognitive learning strategies) that underlie observable changes in behaviors and skills.

Two systems for identifying the outcomes of classroom instruction derive from this perspective: The taxonomies of educational objectives (Bloom, Eglehart, Hill, Furst & Krathwohl, 1984; Harrow, 1969; Krathwohl, Bloom, & Masia, 1964) and learning hierarchies (R. Gagne, 1985). The influence of these two systems on learner assessment has been considerable. We will present an overview of these systems and contrast them with learning assessments based on cognitive learning theory.

Bloom's Taxonomy

Bloom and his associates have developed a system for helping teachers identify the types of learning that they can expect of their students. They refer to their classification schemes as taxonomies of educational objectives. These taxonomies are most helpful to teachers when they are planning their lessons. They help focus teachers on the outcomes or behavioral objectives they want to see students attain as a result of their instruction. The taxonomies also facilitate assessment. Teachers simply need to construct
measurement devices that elicit the same behaviors described in their behavioral objectives.

The taxonomies cover three domains of learning: cognitive, affective, and psychomotor. Let's focus on the Taxonomy of Educational Objectives in the Cognitive Domain (Bloom et al., 1984) to illustrate their focus and purpose.

This taxonomy organizes cognitive outcomes of teaching into six categories:

Knowledge
Comprehension
Application
Analysis
Synthesis
Evaluation

According to Bloom and associates higher level outcomes can only be attained following the learning of lower level outcomes. Analysis (for example, in an American history class) cannot be achieved by learners unless they have previously acquired knowledge of facts, principles, etc., comprehension of those facts and principles, and the ability to apply them. Only then can they be expected to be able to analyze. Likewise, if a learner lacks the skill to analyze empirical results, he cannot be expected to be able to synthesize or evaluate the significance of scientific experiments and research findings.

Knowledge. The knowledge category covers lesson objectives that require learners to remember information like facts, definitions, terms, rules,
words to a poem, etc. The taxonomy includes verbs that describe observable actions which provide evidence that the student has acquired knowledge. Some observable learner actions that suggest knowledge include: defines, labels, selects, states, names.

**Comprehension.** Objectives at the comprehension level require some degree of understanding. Reciting a rule doesn't show understanding. But, restating it in one's own words does, or comparing one rule with another rule would. Some learner actions that suggest understanding are: summarizing, comparing and contrasting, explaining, paraphrasing.

**Application.** Application objectives ask learners to use information that they previously learned in one context in a different context. If a student can borrow in subtraction, reduce a fraction to its least common denominator, or test for the presence of sugar in a chemistry class, and does this with material she has never seen before, the requirements for application have been met. Learner actions that suggest application are: prepare, compute, use, modify, demonstrate.

**Analysis.** Analysis is an activity that requires a student to examine, read, or listen to something and then identify or break it down into its important elements, constituent parts, pointing out relationships and organizing principles. You can analyze an essay into themes, arguments, main ideas, conclusions, etc. Learner actions indicating analysis are: deduce, outline, breakdown, infer, diagram, subdivide.
Synthesis. Synthesis objectives ask learners to create something original or organize something in a unique way. A learner shows that he can synthesize when he discovers a new way to solve a math problem, or design a science experiment. Learner actions that indicate synthesis are: compile, devise, design, create, develop, produce.

Evaluation. Evaluation objectives ask learners to examine a product that they have never seen before and judge its worth according to some set of criteria. For example, after students have learned how to judge an effective political campaign speech, they demonstrate evaluation by critiquing an unfamiliar one. Learner actions indicating evaluation include: appraise, criticize, judge, validate, justify, defend.

Bloom et al.'s *Taxonomy of Educational Objectives in the Cognitive Domain* is a useful reminder to teachers of the variety of ways that students can show simple or complex learning. After you have decided upon the content of a lesson, it is worthwhile to ask yourself about the kinds of learning you want your learners to acquire and the indicators you will use to measure this learning.

But, there are several cautions associated with using taxonomies for deciding what you will teach and assess. First, it might be questioned whether the six levels of the cognitive domain are distinct from one another. While the distinction between knowing something, understanding it, and using it appears obvious, the notion that analysis, synthesis, and evaluation represent distinct and increasingly higher degrees of complex thinking may be problematic. For example, by what basis would you conclude that evaluating a piece of writing is a more rigorous intellectual accomplishment
than creating or analyzing one? The assumption that analysis is a less complex form of thinking than evaluation may persuade some teachers to ignore the former in favor of the latter.

Furthermore, the assumption that lower levels of a taxonomy must be mastered before higher levels can be achieved is unproved and contradicted by some research (White, 1974). While it makes sense that someone cannot read with understanding unless they know the meaning of the words, it is possible for someone to be able to evaluate what they have read and not be able to synthesize a piece of writing or to analyze one. Coaches coach without being expert players at the game. Playwrites write plays without being able to act, and foreman supervise employees without having the same skills of those they supervise. The assumption that higher levels of thinking cannot be expected until lower levels of learning have been acquired may influence teachers to ignore higher levels of learning for certain individuals or classes when in fact, some research suggests that students can learn higher levels of thinking faster than lower levels (Resnick, 1976).

Finally, categorizing behaviors into cognitive, affective, and psychomotor domains does not mean that behaviors listed in one domain are mutually exclusive of those in the other domains. For example, it is not possible to think without having some feeling about what we are thinking, or to feel without some cognition. Also, much thinking involves psychomotor movements and bodily performances that require procedural skills and abilities. Conducting a laboratory experiment requires not only thought but pouring from one test tube to another, safely igniting a Bunsen burner, adjusting a microscope correctly, and so on. It may be convenient for an objective to contain behavior from only the cognitive, affective or psychomotor domain, but one or more behaviors from other domains are
often required for a successful performance. The categorization of desired outcomes into discreet domains may underrepresent the simultaneous importance of multiple outcomes and processes of learning.

Gagne's Learning Hierarchies

Both Gagne and Bloom wanted to end up in the same place: a system for identifying learning goals in terms of their underlying thought process and that organizes them into categories that go from simple to complex. But, they took different roads. Gagne organized learning outcomes into five categories: *verbal information, intellectual skills, cognitive strategies, attitudes*, and *motor skills*. The last two categories are similar to the affective domain and psychomotor domain developed by Bloom and associates. The first three categories encompass learning types that are both similar and different from those included in Bloom's cognitive domain. Let's focus on these three categories.

**Verbal information.** Verbal information is similar to Bloom's knowledge category. It includes knowing the names of explorers, dates, definitions of words, the letters of the alphabet, or the multiplication tables. Although it also includes the ability to state a rule like "When two vowels go walking, the first one does the talking." Verbal information does not assume that the learner understands or can apply the rule when sounding out new words. Instead verbal information provides many of the building blocks for the development of concepts, rules, and generalizations that are covered by Gagne's "intellectual skills" category.
**Intellectual skills.** Intellectual skills cover what the person does with information. Intellectual skills involve making *discriminations* or the ability to recognize one thing as different from another. For example, you make discriminations when you recognize that "b" is different from "d," a horse is different from a mule, pig sounds different from big, or an X-ray of a compound bone fracture looks different than an X-ray of a stress fracture.

The ability to make discriminations is prerequisite for the ability to form *concepts*. A person has concepts when she can label a concrete object (table) or an abstraction (justice) based on specific characteristics or point out that red is different from orange (discrimination). But she doesn't have a concept of red or orange until she can correctly identify all the examples of red as red and not mix them up with examples of orange. Concrete concepts are easier to acquire than abstract or defined concepts. You can teach what a brontosaurus is by showing examples of dinosaurs that are and are not brontosauri. But, the concept of "democracy" requires that you define it for the person using words.

A third level of intellectual skills according to Gagne involves *rule learning*. An example of a general rule is "i" before e except after c." To learn and use this rule, someone has to have already learned the concepts: *i*, *e*, *except*, *after*. So, concept learning is a prerequisite for rule learning. We know that learners have acquired rules when they consistently read certain vowel combinations correctly, solve physics problems that require the applications of Bernoulli's theorem, bisect an angle, or write grammatically correct sentences.

The fourth and highest level of intellectual skills involves *generalizations*. Generalizations are complex or higher order rules like laws of supply and demand, gravitational forces, principles that apply to the rise 3-20
and fall of civilizations, or laws that help explain geological events like earthquakes.

Intellectual skills form learning hierarchies. Any school task like reading for comprehension, solving an algebra equation, writing an essay, or proving a theorem in geometry can be broken down (or task analyzed) into the knowledge, discriminations, concepts, rules, and generalizations that are required to master the task. These skills form a hierarchy in that lower level skills must be learned before higher levels ones can be achieved. Once you correctly classify the task as one involving rule learning, for example, you automatically know that certain concepts, discriminations, and verbal information must first be learned. Figure 3.2 displays a learning hierarchy for conservation skills.

Insert Figure 3.2 about here

**Cognitive strategies.** As you will recall from Chapter 1, cognitive strategies are skills we learn to help us read, write, or solve problems that go beyond the immediate requirements of the task. Although Gagne does not explicitly say so, the learning of cognitive strategies requires the prior learning of certain intellectual skills. For example, learning a strategy to solve four-column subtraction problems involving regrouping and zero's in the second and third columns will require prior learning of certain spatial concepts like "right side," and concepts of number.

Some of the same concerns we had with Bloom et. al.'s taxonomy also apply to Gagne's: assumptions that certain learning is required before other
learning can take place, and ambiguity in the meaning and distinctions among terms like concepts, rules, higher order rules, or generalizations. Moreover, much research fails to support the notion that, for all children, higher levels of learning can only be attained after lower levels have been mastered (Mayer, 1987; Resnick, 1976; Resnick, Siegel & Kresh, 1971).

Evaluation of Bloom's and Gagne's Systems

Bloom and Gagne have made major contributions to our understanding of what is involved in simple and complex forms of learning. Their work has added to our ability to clarify what we mean by learning and to develop valid indicators of learner achievement. They have also underscored the importance of the link between instruction and assessment. But both systems reflect a theory of school learning that derives from a behavioral science tradition, the assumptions of which may not always provide valid explanations of complex human learning (Lieberman, 1992; Hunt, 1995). In particular, this approach and Bloom's and Gagne's taxonomies may view learning as too sequential, hierarchical, and outside the control of the learner to account for some important types of learning.

Research in cognitive science points out that learners are constantly trying to make sense out of the information which they learn and to develop new learnings and understandings in ways that can't be neatly mapped into rigid sequences and hierarchies. Cognitive approaches to school learning, similar to Bloom and Gagne, underscore the importance of information, higher level thinking, and cognitive strategies. But, unlike behavioral conceptions of learning, the cognitive view stresses that learners often derive or construct different meanings from the same information and experiences.
and end up in different places than we would predict from our learning sequences or hierarchies.

In this text, we will take an approach that preserves the strengths of Bloom's and Gagne's systems but acknowledges the "constructive" nature of the learner's contribution to the learning process. In other words we want to maintain the emphasis on clearly specifying what we mean by learning. But, in addition, we want lesson planning to recognize that learning doesn't always proceed in neat, linear steps that require mastery of one thing before moving to another, more complex form of learning.

A Cognitive Learning Model for Determining Outcomes

Let's look in on Lori Copeland who is planning a seventh grade lesson on photosynthesis to see some of the differences in a lesson that has been planned using a cognitive model of learning as opposed to a behavioral model. The main point of Lori's lesson is to help students understand that plants use light to make their own food through a process called photosynthesis. Were she to use a behavioral model, her lesson planing might go something like this:

Let's see. What do I want my students to be able to do at the end of the photosynthesis lesson? We'll, I want them to understand the process of photosynthesis. Understanding is at the comprehension level of the Taxonomy and students can show understanding by explaining photosynthesis in their own words, summarizing the steps in the process using their own words, or predicting what would happen if plants weren't exposed to light. So I'll focus on the following objectives which I can test with a short essay: On an essay test, students will (1) summarize in
their own words the steps involved in the process of photosynthesis; (2) predict and explain what would happen to a plant if it got water and some plant food, but was put in a basement for two weeks where there was very little light.

Now, in order for the students to be able to achieve these objectives, they're going to have to have some knowledge. They'll have to know terms like root system, catalyst, phototropism, process, etc. They'll have to be able to identify the parts of a plant that produce food. They also must memorize in sequence the steps in the process of food making. So, my knowledge objectives in this area will be: Students will (1) recognize the correct meaning of the following words (root system, catalyst, phototropism, stoma, veins, food, energy, and chlorophyll; (2) label the parts of a plant and parts of a leaf; and (3) list the steps in the process of photosynthesis.

I'll start my lesson with activities to help them acquire knowledge for understanding photosynthesis. Once they have all the important concepts, I'll describe the process of photosynthesis. I have a film strip for them to watch and we can examine some plants to point out the structures. Then I'll ask questions to see if they understand what I've told them and we'll have test at the end of class.

Were Lori to plan from a cognitive learning model, her lesson might go something like this:

*At the end of this lesson I want my students to know what it means when we say that plants make their own food. If they*
understand what it means to make their own food, then they should be able to explain why some plants live and some die, why some are strong and some are weak, and they should be able to make predictions about how certain circumstances will affect plant growth.

But, I also want them to do some scientific thinking like raising questions or hypotheses and finding out how to support or disprove them. So, knowing how to ask questions and how to observe and collect information to answer those questions is also an important outcome of this lesson.

Good scientific thinking also involves a willingness to challenge your beliefs about things. So, another goal will be to get learners to reflect on what they know now about how plants make food and to challenge what they know.

I'm sure they all have ideas about what makes plants grow, what is food, how plants get food. If they're like previous students, some of their ideas will be wrong. So, another goal will be to see how their knowledge of things like food, energy, etc. changes over the course of the lesson. They need to see that food means one thing for us and another thing for plants, but there are some things common to each.

I'll start my lesson with a discussion of what is a living thing, what is and is not food, then ask how they think plants get food. From that I'll gradually build to the idea that plants have to make their own food. Then, we can discuss how they do this while I model the scientific way of thinking by my questions. Then, I'll see if their own thinking is becoming increasingly
scientific by observing if they pattern their questions after mine. If I can get them to ask questions that challenge one another's ideas, I will know I'm accomplishing my goal.

At the end of the lesson, I'll develop a test that measures some facts about photosynthesis. But, I'll also ask them to solve some real-life problems pertaining to photosynthesis in the laboratory. I'll ask them to not only solve these problems but also to show me how they arrived at the solution so I can evaluate, not just their answers, but the cognitive process they've used to arrive at their answers.

As you can see, each approach places importance on objectives, and activities to accomplish them. While each model includes the learning of information, the cognitive model also emphasizes the processes by which that information is acquired, qualitative changes in the knowledge base (better organization, etc.) and in problem solving and assessment over the course of the lesson. The behavioral approach measures products, discrete behaviors, answers on a paper-and-pencil test in ways that require that knowledge be taught first, followed by understanding. The cognitive model stresses that learners will arrive at their own understanding if the teacher provides the appropriate context for learning which involves real world examples, giving information, eliciting misconceptions, challenging ideas, modeling problem solving strategies, and giving learners skill in monitoring their own learning. What guides these teaching processes is an analysis of the task the learners are being asked to perform that identifies the necessary knowledge, thought processes, and mental representations that learners must use to think about photosynthesis and similar domains of knowledge.

3-26
Let's return to Figure 3.1 and examine what cognitive learning theory suggests are the important areas of academic learning and techniques for assessing them.

**Declarative knowledge.** The cognitive model of learning stresses that a knowledge base is the foundation of good thinking. This knowledge base consists of declarative knowledge like facts, concepts, rules and generalizations. The goal of a lesson should be not only to increase the size of this knowledge base, but also it's organization so that learners have ready access to it and deep understanding. Assessment of declarative knowledge should be focused on determining not only the accuracy of this knowledge but also how the various facts, concepts, and generalizations are related to one another, and how this organization changes over time.

To measure declarative knowledge, teachers can use traditional pre and post paper-and-pencil tests but also assess changes in its organization using techniques like "ordered trees," "conceptual maps," and "story grammars" which you will learn about in Chapter 7. The ability to problem solve using ones knowledge of facts, concepts, rules and generalizations can be assessed by traditional essay questions (which will be presented in Chapter 6) and also by performance tests that will be treated in Chapter 7.

**Procedural knowledge.** Knowledge also involves procedural knowledge which is knowing how to do something. A procedural knowledge base develops from initial states where learners practice a sequence of skill components to a stage where the procedures are used automatically and intuitively. In addition to traditional tests that ask learners to show their work, procedural knowledge and its development can be assessed with
checklists and other observation techniques which you will learn about in Chapter 8.

**Cognitive strategies.** As pointed out above, the cognitive learning model emphasizes not only the products of learning but also the processes that produce it. Consequently, cognitive learning strategies should be taught and assessed. The goal of such instruction should be the flexible and adaptive use of such strategies as the problem context demands. In Chapter 9 we will present examples of cognitive learning strategies in a variety of areas with suggestions for how to assess their efficient and flexible use with rating scales, debriefing interviews, learning logs and journals, student self-evaluations and "think alouds."

**Metacognition.** Learning a cognitive strategy doesn't always result in the use of that strategy. Learners must be taught when and where to use strategies and how to monitor the learning that results. Chapter 9 will present techniques for assessing how well learners regulate their own learning.

**Motivational orientations.** In Chapter 2 we discussed the role of motivation in learning and its link with teaching and testing. We discussed how attributions, self-efficacy, and goal focus affect effort. Chapter 10 we will present methods for assessing motivation and its development using such techniques as attitude scales and observations that measure achievement motivation.

In Chapter 11 we will discuss grading from a cognitive learning framework and we will offer ideas on how to communicate assessment results to learners and their parents. Now, let's turn to some ideas and 3-28
techniques for assuring that your assessment techniques meet standards of validity, reliability and fairness.

Summary
This chapter has shown you how to establish an assessment mission and determine learning outcomes. Establishing your assessment mission involves choosing a norm-referenced, criterion-referenced, or growth referenced frame of reference and establishing how you will set expectations, make a diagnosis, monitor learning and assign grades. This chapter also introduced Bloom et al's Taxonomy for the Cognitive Domain and Gagne's learning hierarchies to help you plan an authentic learning assessment. Finally, this chapter defined five important areas of academic learning: declarative knowledge, procedural knowledge, cognitive strategies, metacognition, and motivation orientation. We will revisit each of these important areas of learning in the chapters ahead.
Activities

1. Using Figure 3.1 as a guide, write a mission statement for an authentic learning assessment program in your classroom.

2. Give examples of tests that either you have taken or know of that were (a) norm-referenced, (b) criterion-referenced, and (c) growth-referenced. In your opinion, what is an advantage and a disadvantage of each of these frames of reference?

3. How would you answer someone who said "You can't be a good coach unless you've been a good player, but you can be a good teacher without having been a good student." Explain whether you agree or disagree using either Bloom's or Gagne's concept of learning.

4. Using the two lesson plans of Lori Copeland which you read in this chapter, write two new scenarios of a lesson plan you will be teaching, one representing the behavioral (ABC) model of learning and the other the cognitive model. Which model do you believe will be most effective for achieving the goals of your lesson?

5. For the two scenarios you have written, identify any declarative knowledge, procedural knowledge, cognitive strategies, and motivational orientations that will be necessary for achieving the goal of your lesson.

Suggested Reading


This book discusses ways to categorize learning outcomes using Bloom's, Gagne's, and other cognitive learning categories.
<table>
<thead>
<tr>
<th>Learning Mechanisms</th>
<th>Initial States (Novice)</th>
<th>Transitions</th>
<th>Desired States (Expert)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declarative Knowledge</td>
<td>Naive Theories → Recapitulation → Deep Understanding and Misconceptions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Procedural Skills</td>
<td>Skill Components → Coordination → Efficient Intuitive Use</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive Learning Strategies</td>
<td>Selective Attention → Internalization → Multiple Flexible Strategies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monitoring/self Regulatory Functions</td>
<td>Action Orientation → Engagement → Adaptive Action Control</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivational Orientations</td>
<td>Mastery Orientation → Effort → Achievement Motivation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chapter 4
Quality Assurance

The initial chapters of this book have emphasized that teaching and assessment coexist in dynamic interaction. They have the same mission, to enhance student learning, and the same goals, to develop knowledge, thinking skills, and motivation. Teachers set goals and objectives that assessment measures and evaluates. And, assessment informs teachers about learner progress and the usefulness of their teaching methods.

This chapter introduces some concepts that will assure that your assessment system accomplishes its mission. It addresses two related questions:

1. How do you know that your assessment system is measuring what you want, and
2. How do you know that your assessments of student learning are free of error?

The first question relates to a standard called validity. You may develop an assessment activity to evaluate how well a student can think critically. But, how do you know that it's measuring critical thinking and not memorization?

A standard called reliability is the focus of our second question. How do you know that your conclusions about a learner's ability to think critically, based on observations and questions asked during Wednesday's lesson, would be the same as those based on observations and questions during Friday's lesson?

Quality assurance is a term we will use for building validity and reliability into your authentic learning assessment system. In this chapter we
will identify indicators of validity and reliability and suggest methods of obtaining them.

Validity

In Chapter 3 you learned about different approaches to determining the outcomes of your learning assessment system. The outcomes typically referred to "inside the head" phenomena like "knowledge," "organization of knowledge," "cognitive strategies," "problem solving," and "metacognition." You can't see any of these. So, you need to identify observable actions that suggest that they are being learned. A "test" is a vehicle for measuring these actions. A test can consist of a written sample of your students' work, observations of their behavior, an interview in which they express their thoughts and opinions, a performance assessment of their skills or a product representing what they have learned. In other words, validity concerns how well your "tests" measure what you want your students to learn.

Validity encompasses three important questions. The first two are:

1. Am I assessing the behaviors that I want my students to learn?
2. Am I assessing what I teach?

The first question measures a type of validity called construct validity while the second measures instructional validity.

A third dimension of validity concerns the purpose of your assessment. For example, do you want your assessment system primarily to promote learning, grade students, or improve your teaching. Thus, the third validity question is:

3. Does my assessment system accomplish its mission?
Let's examine each of these questions and point out what you can do to assure that these types of validity are a part of every assessment you make.

Construct Validity

The following essay appeared on an 11th grade American History test:

Imagine that it is 1938 and you are an citizen living in Ohio. Because you are interested in economics and concerned about the country's future, you try to keep yourself well-informed. You have just heard "replays" of Franklin Roosevelt's and Hewy Long's speeches on the radio. Your cousin, who has not listened to the speeches, asks you about some of the problems facing the nation.

Write an essay in which you explain the most important issues your cousin should understand. Your essay should be based on three major sources: (1) the specific facts you know about American history, especially what you know about the history of the Depression, (2) the Roosevelt and Long speeches, and (3) what you learned from the supplemental readings.

Be sure to show the relationship among your ideas and facts.

This essay was designed to assess an understanding of American History: the acquisition, organization, interpretation, and application of specific content knowledge in new contexts (Baker, 1994). From reading the question it appears to require these thought processes of learners. So the answer to the question, "Does this assessment measure the qualities of thinking that the teacher intended?" seems to deserve a "yes."

But does it? Suppose that the two-week unit on the Depression that preceded this assessment included a review and discussion of Long's and Roosevelt's speeches, an analysis of the supplemental readings, and class exchanges about the most important ideas and issues prevalent during that
period in American History. If this were the context in which the essay question were asked, then all a learner need do is to recall what was said in class and parrot that back during the exam. Under these circumstances, the answer demands mostly memorization, which is a far cry from the understanding that the question was designed to measure. In other words, under these circumstances the task lacks construct validity.

An assessment has construct validity when it produces learner behaviors that bear a direct link to the cognitive activity you want to assess. If you want to determine how well one acquires new knowledge and applies this to solve a problem, then you must design an assessment context that requires those cognitive processes. The assessment context is more than the written question. It involves prior exposure to test content, limits placed upon the learners response (called test constraints) and the nature of the response being required. In this example, the way the test question was written suggests it has construct validity for assessing understanding. But this understanding also depends on the cognitive processes actually required by the learner to answer the question correctly.

Looks can be deceiving. A test that appears to be assessing one type of cognitive learning, like analysis or application, may actually require a different type of cognition. By the same token, asking a learner to list the principal reasons for the Depression may tell you little about whether the learner understands them. Likewise, labeling the states on a map is no indication of whether someone knows how to travel from North Dakota to Iowa.

Construct validity is more difficult to achieve as the cognitions you want to assess become more complex. Assessing whether a learner has basic knowledge of people, places, dates, and events about the Depression may be
relatively easy. Determining how organized and related the learner's information base about the Depression is may be more demanding. And, assessing whether the learner can draw upon this knowledge to solve important problems may be a challenge to any teacher's test-making ability.

Assuring Construct Validity

The basic challenge in achieving construct validity is assuring that the behaviors you decide to observe and measure are valid indicators of what a student knows and thinks. Here are some important steps for doing this.

**Step 1: Specify the construct domain.** This is the first and most essential part of assuring construct validity. It requires that, as best as possible, you delineate what you mean by the type of cognitive activity you want to assess. For the question about the Depression, the teacher wanted to assess understanding which would require the following cognitive activities: knowledge acquisition, knowledge organization, interpretation, and application of specific content knowledge in new contexts.

The more indicators of the behavior you want to measure the better. You can consult Bloom's or Gagne's Taxonomies in Chapter 3 to see what is meant by knowledge, comprehension, analysis, concept formation, problem solving, etc. And, don't forget to consider the thought processes of experts when they are engaged in problem solving, writing an article, solving a math problem, interpreting historical events, designing an experiment, etc. to see how well your test questions match theirs.

Some cognitive outcomes are more difficult to specify or delineate than others. If you want to assess content specific knowledge, then asking students to recall facts, recognize important dates, or give definitions can be
assessed by paper and pencil tests. If you want to assess whether your students know how to use a certain procedure, then specify the steps of the procedure and perhaps use a checklist to see how well it is followed. With constructs like "making connections," "becoming a mathematical problem-solver," "communicating complex ideas," "deep understanding," specifying what you mean can be a more challenging task.

**Step 2: Choose as many indicators of the domain as possible.**

When specifying the construct domain, identify all the behaviors that represent it. For example, let's say that you are a 4th grade math teacher and you want to assess understanding in the area of fractions. You decide that understanding involves cognitions like problem solving, reasoning, and communication. You break this down further into seeing mathematical relations, organizing information, using and discovering procedures, formulating hypotheses, evaluating the reasonableness of answers, justifying an answer or procedure, and explaining your solutions to someone else.

The best way to assure construct validity for assessing deep understanding is to design a task that requires as many of these indicators as possible. Developing a fraction assessment task that only requires using a procedure and justifying an answer would not have as much construct validity for assessing understanding as one requiring three or four indicators.

Notice the number of indicators of deep understanding in the essay question about the Depression: (1) acquisition of knowledge, (2) organization of knowledge, (3) interpretation, and (4) application. All four of these indicators could be rated or judged from the students' writing.

Elementary school teachers who wish to assess reading comprehension can ask their learners to do several of the following: summarize what was

4-6
read in their own words, predict what would happen next, describe the main idea of the passage, relate the surprise ending to an event in their own life. All of these represent multiple indicators of the cognitive activity being measured. Don't forget Bloom's *Taxonomy of Educational Objectives*, which we studied in Chapter 3, for indicators of thinking skills.

**Step 3: Design the assessment task so that it requires only the cognitive skills relevant to the construct you wish to assess.** Sometimes when we design an assessment task we inadvertently measure skills we didn't intend, or require that students use skills that are not relevant to the construct which we plan to measure. For example, the question about the Depression was designed to measure understanding, which was defined as acquisition, organization, interpretation, and application of knowledge. But what if the assessment task included reading material that was too difficult for many learners to understand? In that case, the assessment would not measure the construct of understanding, but rather reading comprehension.

Using the same example, if the scoring standards for this essay gave more importance to *how* the student wrote than *what* the student wrote, this question would measure writing more than thinking ability. Relevancy is judged against the purpose of the assessment which you identified in Step 1.

Remember to design an assessment context that requires construct relevant skills and avoid contexts that make the assessment tasks unrelated to your instruction. After you have designed the assessment task ask yourself, or a colleague, whether it requires thought processes or constructs irrelevant to your instruction like reading ability, writing ability, memory, or background knowledge that gives unfair advantage to learners who happen to have had particular experiences.
Step 4: Write task directions that require the thought processes that you intend to assess. You should phrase the task in such a way that the desired thought processes underlying the response are attained. You should select precise terms and phrases that communicate to the learner what you want them to do. Notice how the essay question about the Depression was phrased. It used specific words and phrases to communicate to the learner what she or he was expected to do: "explain the most important issues...", base your essay on "what you learned from supplemental readings," and "show relationships among your ideas and facts."

When writing directions we sometimes use words and expressions that are ambiguous to learners, however clear they may be to us. Here are some ambiguous words and expressions: integrate your findings, synthesize the information, analyze the issues, evaluate the findings, interpret the passage, formulate an hypothesis, etc. Each of these expressions needs some clarification that can accompany the test or during the lessons that precede it, so that students know what the directions require of them. The best way to check on the clarity of your phrasing is to write a model answer or set of guidelines to evaluate the answer and see whether you did what the task specified. This usually results in some revision of the directions and task specifications.

APPLICATION
Write an essay question that measures "deep-understanding" of a topic you will be teaching following steps 1-4 above. In the space below tell what you did in planning your essay questioner to accomplish each of the four steps.

1. Specify the construct domain:

2. Choose indicators of the domain:

3. Include only relevant cognitive skills:

4. Write task directions that require the relevant thought processes:
Gathering Evidence of Construct Validity

If your tests have construct validity, then after you have rated or scored them, and have come to conclusions about what your students know, you will observe the following:

(1) The students who do best should be among those learners you know to be more knowledgeable and expert. You have lots of prior information about your learners' knowledge and background. This will come from previous grades, listening to their answers in class and from reading their homework and classwork. All other things being equal, those whom you judge to know more in a particular content area should do better on your assessment. If not, it may indicate that your assessment is unrelated to your instruction.

(2) Learners should do significantly better on assessments given at the end of a lesson or unit than on similar assessments given at the start of the unit. In other words, one bit of evidence of the construct validity of an assessment is whether learners who have had instruction perform better than learners who have not.
(3) Other assessment data that measure similar abilities or constructs should agree with the conclusions about learners that you make from your assessment. For example, students take standardized achievement tests of reading and math. All things being equal, students who show strong reading comprehension or math problem solving skills on these tests should also be among those who do well on your tests that measure the same skills.

Instructional Validity

In Chapter 2 we maintained that the link between assessment and teaching is a principal influence on academic motivation. There is a type of test validity that relates to this link. We refer to this type of validity as instructional validity.

While most teachers are certain that they test what they teach, sometimes their learners are not as convinced. One of the most common complaints of students whether in grade school or college is that their tests do not always measure what they were taught (Tuckman, 1988). When there is a tenuous link between testing and teaching, two problems occur: students who have learned what was taught are unable to demonstrate it; and they are evaluated over topics for which they have received little or no instruction.

Thus, teachers often fall into two traps when designing assessments:

(1) They test content areas or skills that they didn't teach. This often occurs when teachers hold their students accountable for textbook content or skills that were not discussed in class, assigned for homework, or encountered in workbook exercises. Sometimes it may be thought that, although some areas were not covered, "good students will learn it anyway."
(2) The assessment places more emphasis on certain domains of knowledge or skills than was reflected during instruction. A common complaint of students is that a test covered too many areas that were only briefly discussed in class and ignored much of what the teacher presented during lessons.

Thus, an assessment has instructional validity when it asks learners to do what was taught during their lessons and with the same degree of emphasis. If students saw a teacher demonstrate how to reduce fractions to their least common denominator, practiced this, and solved similar problems for homework, than an assessment that is valid for assessing this skill should ask them to reduce fractions and not to transform them into decimals.

Assuring Instructional Validity

An assessment with instructional fidelity meets two standards:

(1) It reflects the goals and objectives of your lessons.

(2) It gives the same emphasis to your goals and objectives as did your lessons.

One way to insure instructional validity is to construct an assessment blueprint. Table 4.1 shows a teacher's assessment blueprint for a three week unit on how the earth's surface changes. The content or topics covered are listed down the first column and the cognitive learning outcomes emphasized in class are identified across the top. This particular unit emphasizes memory or accumulation of domain specific knowledge, the use of certain procedures and cognitive strategies, and problem solving ability. The percentages indicate the teacher's emphasis for both content and cognitive behaviors. The cells identify assessment techniques that measure the intent of the lesson and
the percentages reflect how much weight or importance each technique has in the overall evaluation or grade. The totals for rows and columns reflect the emphasis (in terms of percentages) that the assessment must have to match the instructional emphasis of the class.

Insert Table 4.1 about here

There is a variety of systems for constructing assessment blueprints (Kubiszyn & Borich, 1996; Oosterhof, 1996; Gronlund, 1993). Ideally they should be constructed when you plan your lessons and complete your task analyses. That way what you teach and its emphasis are easily remembered.

When you design your assessment, remember to follow this procedure.

1. Have the blueprint in front of you. Do not rely on memory to tell you if you emphasized knowledge acquisition or problem solving and to what degree.

2. Write your assessment questions, tasks, and specifications from a learner's perspective. Ask yourself what information and cognitive skills are required to do these tasks and did you prepare your learners for them?

3. For each assessment task that you develop, refer to the cognitive learning outcomes that it is intended to assess. Ask yourself: "Does this question reflect the level of cognitive complexity indicated in the learning outcome?"

APPLICATION
Think of two tests that you have taken—one which you felt had instructional validity and another which you felt did not. Identify the tests and list the distinguishing features of each test that caused one to have instructional validity and the other to lack it.

Distinguishing characteristics of the instructionally valid test:

---

---

---

---

---

Distinguishing characteristics of the instructionally invalid test:

---

---

---

---

---

In college teaching most of the evidence for instructional validity comes from comments that students make on course evaluation surveys.
Specific questions address the relevance of test content to instructional objectives. But, in the elementary and secondary school evidence for this type of validity comes to your attention informally: you overhear students griping about how unfair the test is or they tell it to you directly.

Take these comments seriously. They suggest that you should routinely assess the instructional validity of your assessment by asking learners, at the end of a test, to anonymously agree or disagree with statements, such as the following:

Strongly Agree    Agree Not sure    Disagree    Strongly Disagree

☐       ☐       ☐       ☐       ☐       ☐       ☐

Class and text adequately prepared me for the test.

☐       ☐       ☐       ☐       ☐       ☐       ☐

A lot of questions on the test came as a real surprise.

☐       ☐       ☐       ☐       ☐       ☐       ☐

Too many questions asked about things we didn't cover in class.

☐       ☐       ☐       ☐       ☐       ☐       ☐

Overall, this test was fair.

Consequence Validity

Until recently, discussions about the importance and meaning of test validity focused primarily on the types of validity discussed above. But, many assessment experts now believe that validity considerations should also include the use made of tests scores and the consequences they have on
teaching (Messick, 1989; Gipp, 1995). In other words, validity also encompasses whether the learning assessment system that you designed changes both your and your students behavior.

If the purpose of your test is to facilitate learning and motivation, then consequence validity requires evidence of this. If a significant number of learners have not become better motivated, increased the size and complexity of their knowledge base, or learned how to think better, then your assessment lacks evidence of this type of validity. If you want your assessment to be used to help you adjust your teaching methods to become a better teacher, then there should be some evidence that your teaching changed as well.

The best way to assure consequence validity is to have a plan that specifically describes how you will use the results of your assessments. This plan should:

1. Describe how you intend the results of your assessment to be used? Check all that apply:

   - Grading
   - Improving teaching
   - Improving learning
   - Monitoring learner development
   - Enhancing motivation
   - Informing parents

2. For each purpose, identify what evidence you will gather to show that the purpose is being achieved, for example:

   - Percentages of answers correct
☐ Changes in test scores across the semester
☐ Opinion surveys of learners and parents
☐ Motivational assessment
☐ Ratings of teaching effectiveness and improvement

(3) For each purpose, describe how you will use the evidence to achieve it.

Reliability

Assuring validity means that your assessment methods are measuring what you want them to measure. But, this is only one aspect of quality assurance. Another concerns the accuracy of the scores, grades, ratings, or judgments that you make about a learner from your tests. You can make a paper and pencil test, or construct a performance task, that has the potential to measure what you want but which may not reliably measure your learner's behavior.

For example, take the essay question about the Depression presented earlier. The question appears valid for assessing understanding. But, after the students answer the question, the teacher has to read their work and make judgments about the degree to which they reflect "understanding." If two teachers read the same essay and assigned some grade to it without consulting one another, would they agree? Reliability is the extent to which your judgments about a learner's test or task performance are free from bias or subjectivity.

Reasons Why Assessments Lack Reliability
Here are some assessment situations. As you read them, reflect on how comfortable you would feel if you were you the teacher described.

A fifth grade teacher developed a five question true-false test to determine the factual knowledge of her learners in U.S. geography. The questions covered four chapters. Sean got a "C" because he missed two questions. He told the teacher that he knows a lot more than a "C" but the questions on the test weren't enough to reflect what he knows. The test just asked questions over the areas in which he was the weakest.

Rose, a second grader, got an "unsatisfactory" on her six-week report card for "works cooperatively with others." Her mother questioned the teacher about why. The teacher explained that she based her judgment on how Rose behaved during the cooperative learning groups which met for 45 minutes every Friday. The parent explained that Rose was absent for all but two of these groups and wondered whether the teacher saw enough of Rose to come to such a judgment.

Ozzie, a ninth grader, was upset about getting only 13 of 20 points on one of the essay questions on his American history test. The teacher explained that he failed to discuss the issue of illness as a factor in the Lewis and Clark expedition. Ozzie said he knew about that but didn't know he was supposed to have it as a part of his answer. The essay question read: "What do you consider to be the principal factors contributing to the successes and failures of the Lewis and Clark expedition?"

Mr. Snead, an eleventh grade American Literature teacher, used the following method to grade his student's book reviews of Hemingway's *For Whom The Bell Tolls*: They would be graded on an A to F basis with +s and -s. After reading an essay he would make a global judgment about how good it was and put it in one of six piles corresponding to the letter grade. Then he would go to each pile and assign -s or +s. Laura got a "C-" and wanted to know exactly what was the difference between a "C-" and a "C" and what she needed to do to get a B.
Each of these vignettes portrays a reason why assessment judgments made by a teacher can be unreliable: limited sample of behavior, small number of observation occasions, unclear tasks, and scoring imprecision. Let's consider each of them.

**Limited sample of behavior.** There are many practical reasons why assessments have to adhere to certain time constraints. Thus, the number of questions asked on an assessment have to be limited. But, this comes at a cost. The fewer questions or tasks that you observe, the more likely it is that your judgments of the learner based on those tasks will be unreliable. Imagine trying to make a judgment of someone's dart-throwing ability based on one toss at a target. Maybe a finger slipped, the person sneezed, someone bumped into her, a clap of thunder startled her and shook the target?

The same holds true during an academic assessment whether that assessment be a test, an observation, or a project. At any given time there are random influences affecting the learner. The more the learner is asked to do, the more likely you rule out these random influences on performance. With only a five question quiz, it is likely that random influences like momentary lapses of attention, anxiety, a classroom disturbance, or having asked several questions over areas the student forgot to study or for which he wasn't in class will affect performance. But, with a 30 question test, these factors play a less prominent role.

**Small number of observation occasions.** Learners have good and bad days just like teachers. A test, a one time observation, a single opportunity to do a science procedure, or a no revision policy for a writing assignment, increases the likelihood that factors other than the learner's
knowledge, thinking skills, or writing ability will affect the performance that you see.

Judgments about non-cognitive attributes like motivation, attitude or getting along with others are especially susceptible to unreliability due to limited opportunities to observe. Thus, the more occasions on which you base your judgments, the more reliable they will be.

**Unclear tasks.** Lack of clarity or ambiguity is a major hazard to reliability. If a learner performs poorly on a test of knowledge, it should be because he lacks knowledge and not because he did not understand the questions. Likewise, if a student fails to show evidence of problem solving during a science experiment it should be due to her lack of this cognitive skill not because she didn't understand what was expected of her.

Vague questions, ambiguous requirements, or unclear directions allow learners to interpret the task in ways that you did not intend. In such cases your judgments about what is lacking or missing in the learner's performance may say more about your test development skills than the learner's cognitive abilities.

**Scoring imprecision.** Scoring is a significant contributor to unreliability, especially for essay questions, writing assignments, history projects, or science experiments that are designed to assess complex thinking or deep understanding. Vague scoring standards increase the likelihood that a teacher's judgment will be influenced by factors unrelated to what the students actually wrote, said, or did.

For example, the American history question about the depression was designed to assess knowledge acquisition, organization, interpretation, and
application. But, without a detailed scoring plan to assess these qualities of thinking, a teacher could easily be influenced by writing style over the substance of what was written. Moreover, lack of a scoring plan may cause factors like teacher fatigue to play a prominent role in the measurement process.

Assuring Reliability

Certain assessment goals present a greater challenge to validity than others. This also is true for reliability. Building reliability into a project or test to assess problem solving ability and critical thinking presents a greater challenge to your assessment design capabilities than constructing a paper-and-pencil test to assess the accuracy of factual knowledge. Still, regardless of the learning outcomes you wish to assess, there are some general strategies that you can use to assure reliability in any assessment context: Increasing the number of performances to be scored or rated, increasing the number of occasions on which to score or rate, writing clear task specifications, and increasing scoring objectivity. Let’s take a closer look at each of these.

**Increasing the number of performances.** If you want to find out how the parents of students in your school feel about homework, and you could not ask them all, you would be concerned about two issues of relevance to reliability: number of parents sampled and their representativeness. You want to make sure that the parents you survey aren’t from one segment of the community and that you have enough parents to adequately represent all segments.

The same considerations apply when designing an assessment. You can't ask questions over all the content; so you ask enough questions to
adequately represent the most important aspects. If you are trying to assess knowledge of physics to determine misconceptions you need to identify the most important misconceptions that learners have about physics, randomly sample from this domain of misconceptions, and write true-false or multiple choice questions over them. If there are 50 misconceptions that learners have about forces, motion, gravity, magnets, electricity, etc., randomly sample 15 of them and put them into question form. Someone who endorses 90% of the misconceptions would probably have endorsed 45 of the 50 misconceptions had you included them all on the test.

Thus, you need to do three things to satisfy this standard for reliability pertaining to assessment length: (1) define the domain of important knowledge, (2) select an appropriate number of bits of knowledge or information from this domain, and then, (3) randomly sample from this domain.

The difficulty of doing this comes when you are trying to assess domains other than simple knowledge like understanding, problem-solving or the skilled use of cognitive strategies. Designing more than one or two tasks to assess these areas would take too long. The Depression essay question may require a class period to be answered adequately. So, in situations where you are designing an assessment to measure complex skills, make sure that you observe or rate the skills on more than one occasion.

**Increasing the number of occasions.** When you conclude that a learner is a good problem-solver in math fractions, or a strategic thinker in biology, or an effective communicator in writing based on one observation, or one essay question, or one short story, a reasonable question to ask would be: Would you come to the same conclusion about this person's
understanding, problem solving, or writing ability if you were to observe them on a different but related task on another day?

The only way to assure that your judgments based on one observation generalize to other occasions is to plan other opportunities to observe. Sometimes this occurs naturally, for example when evaluating a piece of writing where multiple drafts are turned in for feedback from the teacher. In this situation, the reliability of the teacher's judgments or ratings about a written work is increased due to the multiple opportunities to rate it. The reliability of scoring science or social studies projects, complex math problem solving, or oral presentations will be enhanced by giving learners multiple opportunities to present and improve their work.

When the assessment is of such a nature that repeated measurements of the same task are not practical or feasible, then it is important to plan for several observations on different occasions of math problem solving, answering open-ended essay questions, using particular scientific procedures or methods, or designing an experiment. In such cases the tasks and context should be different but assess the same underlying construct like acquisition and communication of knowledge, or application of scientific procedures.

**Writing clear task specifications.** The key consideration here is reducing task ambiguity and this can only be accomplished by writing clear, and concise questions, tasks, and directions. In the following chapters, we will show examples of how assessments can be written so that learners will know what they are supposed to do. We will offer suggestions for improving the clarity of both paper and pencil and performance tests.
Increasing scoring objectivity. Some assessment formats present a greater challenge to scoring objectivity than others. Essay questions, like the one described above, are more difficult to score reliably than true-false or multiple-choice questions. But, even in the latter cases, scoring keys are important.

Essay questions and performance assessments can be scored in a variety of ways, which include holistic techniques, rating scales, and analytic scoring procedures each of which has its advantages and disadvantages that will be described in subsequent chapters. But, whichever method you choose, you should design a scoring plan--called a rubric--that defines the level of accomplishment you desire for products (poems, essays, drawings, maps) complex cognitive processes (skills in acquiring, organizing, and using knowledge), use of procedures (physical movements, use of specialized equipment, oral presentations), and attitudes and social skills (habits of mind, cooperative group work).

Summary: A Quality Assurance Checklist For An Assessment System

Your learning assessment system will accomplish its mission if it meets standards of validity and reliability. The checklist below summarizes the various aspects of these two considerations. Review it before administering each test, observation, or performance rating system that you design.

QUALITY ASSURANCE CHECKLIST

I. VALIDITY
   A. Construct Validity
have you identified the construct domain?

have you specified a number of indicators of this domain?

have you chosen several of these indicators for which you will design tasks, such as:

(1) writing skill
(2) memorization
(3) reading ability
(4) prior knowledge

have you constructed tasks that assess only those indicators?

have you asked a colleague for his/her opinion on what skills the tasks require?

have you taken steps to make sure that the task directions require the thought processes that you want to assess, such as:

(1) Model answers
(2) Review by colleagues

B. Instructional Validity

have you defined the goals and objectives of your lesson?

do your assessment tasks measure only those goals and objectives?

have you estimated the extent of class time devoted to those goals and objectives?

does the relative importance assigned to each assessment task match the extent of instructional time devoted to them?

have you developed an Assessment Blueprint?
C. Consequence Validity

☐ Have you identified the uses to which your assessment results will be put, for example:
   (1) grading
   (2) improve teaching
   (3) improve learning
   (4) monitor development
   (5) enhance motivation
   (6) inform parents
   (7) other ____________

☐ What steps will you take to assure that your assessment results accomplish the purpose(s) identified above?

II. RELIABILITY

☐ Have you adequately defined the domain of knowledge or skills that you wish to assess?

☐ Have you included sufficient numbers of tasks or questions that adequately represent or sample this domain?

☐ Are you observing the learner on more than one occasion while he/she is displaying the learning?

☐ Do you observe more for complex than simple learning?

☐ Have you taken steps to assure that your tasks, questions, and directions are unambiguous?

☐ Have you developed a scoring system for the task, which may include:
   (1) Answer key for objective-type questions
   (2) Essay questions/performance assessments
(3) Holistic scoring, if so:
   ☐ Have scoring rubrics been defined?
   ☐ Have you defined each category or level?

(4) Rating scales, if so:
   ☐ Have you defined the most important characteristics that show a high degree of the trait?
   ☐ Have you identified common errors or mistakes which are most justifiable for achieving a lower score?

(5) Checklists, if so:
   ☐ Do the behaviors on the checklist occur frequently enough to be observed?
   ☐ Can each behavior be rated as present or absent?

(6) Analytic scoring, if so:
   ☐ Has a model response been developed?
   ☐ Are points assigned to different aspects of this response?
   ☐ Does the weight assigned to each aspect match your instructional emphasis?

Activities
1. Explain the difference between the reliability, construct validity, and instructional validity of a test. Now imagine that you just discovered that a test you gave lacked each of these. What could you do to improve each before you gave the test the next time?
2. For the test above, explain how you would show evidence of (a) reliability, (b) construct validity, and (c) instructional validity after you made the changes you have indicated?

3. Using Figure 4.6 as your guide, construct an assessment blueprint for a unit test in your subject or grade. Provide percentages that indicate your intended emphasis for both subject-matter content and learning outcomes and be sure the total percentages for rows and columns reflect your instructional emphasis.

4. Show that the test you used for Questions 1 and 2 has consequence validity by describing (a) how you intend to use the results of this test and (b) what evidence you will gather to indicate that your purpose for this test has been achieved.

5. Describe a test you would like to develop by using the Quality Assurance Checklist at the end of the chapter as your guide. Provide as much detail in your description as the Checklist calls for.

Suggested Reading


An excellent discussion of traditional psychometric notions of reliability and validity. But it also discusses these notions as they apply to current concerns about the usefulness of educational assessment.
Chapter 5
Assessing The Knowledge Base

One of the principal criticisms of American education is that learners spend too much time memorizing information and too little time learning to use it (Sizer, 1984). These criticisms have had an impact on teacher training, curriculum, and testing. There is a much greater emphasis now than before on such outcomes as critical thinking, learning to learn skills, and deep understanding.

But these worthy efforts to increase the importance that teachers assign to thinking have often led them to underrate the significance of knowledge (Bruer, 1993). And, one of the most important findings of cognitive learning theorists is that knowledge is indispensable for good thinking. What has convinced them of this? Let's look into an influential study about the importance of knowledge for complex thinking.

Knowledge Versus IQ

Probably most people believe that to be a good thinker, to solve problems like an expert you have to be smart—you have to have a high IQ. In 1986 Stephen Ceci of Cornell University conducted research on just this issue. He wanted to know how important IQ was for complex thinking or problem solving in comparison to knowledge. Or, in other words, if you had to place your bet on who was a better problem-solver in a specific area of expertise, would it be the person with the high IQ but little knowledge or the person with average IQ and a lot of knowledge? Who would you choose?

The specific area of expertise that Ceci studied was horse race handicapping. An expert handicapper has to consider many things in deciding
which horses in a race are likely winners. They have to know statistics pertaining to the jockey, horse, and track—like lifetime earnings, lifetime speed, racetrack size, track surface conditions, purse size, etc. Fourteen variables in all are examined for each horse in the race. An expert has to know about all these statistics and engage in a complex reasoning process to decide which horse is likely to win, place, or show.

In the Spring of 1982 Ceci went to the Brandywine Raceway in North Wilmington, Delaware and approached 110 people about being in his three year study. He settled on 30 men who were avid racetrack patrons. Fourteen of them were classified as experts based on their ability to pick the top horse nine out of ten races and the top three horses at least five out of ten races. The sixteen nonexperts were no where near as capable of picking the top horses.

Ceci took great pains to make sure that the only difference between the experts and nonexperts was handicapping ability. Both groups had similar IQ's (99.3 versus 100.8), occupational prestige, average years of education (10.1 versus 10.2) and average years of track experience (17.4 for nonexperts, and 15.1 for the experts).

Then, he gave both groups an array of statistics about 50 horses. Each horse was pitted against a "standard horse" in 50 two horse races. The task was to pick the winner. As expected the experts did much better at picking winners than the nonexperts and displayed much more sophisticated reasoning. In fact, their handicapping ability and reasoning matched that of a professional handicapper included in the study.

What was unexpected was the relationship between reasoning ability and IQ. IQ had nothing to do with picking winners or complex reasoning. Low IQ experts, in some cases with IQ's as low as 81, 82, and 84, clearly
surpassed high IQ non-experts, some with IQ's of 113, 118, 130. What the former group had that the latter group lacked was knowledge about how to handicap. The experts knew exactly which statistics were important and which were not and the relationship among all these facts. Ceci's conclusion: in a specific area of expertise, IQ is unrelated to good thinking. Knowledge, information, facts and how to use them in decision making separate the skilled performer from the unskilled.

Does this conclusion hold for things like reading comprehension in school-age children? A group of German researchers have provided some convincing evidence that it does (Schneider, Korkel, and Weinert, 1989). They worked with children in grades 3, 5, and 7. They identified those who had a lot of information about topics like soccer and those who did not but were comparable in IQ. The learners with more knowledge were much more capable of detecting inconsistencies and contradictions in reading passages about soccer than learners with less knowledge regardless of IQ.

In this country Walker (1987) showed that low IQ learners who knew a lot about baseball got more out of a baseball passage than high-IQ learners who knew little about the sport. Hall and Edmondson (1992) found similar results with basketball expertise. What all these studies point to is that those of your learners with substantive knowledge about a specific subject will be your best thinkers. Those who know little will not reason well regardless of their IQ's. The learners who get most out of reading a chapter on the Depression, weathering, photosynthesis, or the legislative process will be those who have the most information at the time they are reading.

The Rightstart program (Griffin and Case, 1975) referred to in Chapter 1 makes a similar point. Their work shows that low SES learners are at risk for low math achievement because they come to first grade without particular
math information not because they have lower aptitude for math. In the area of science instruction, Hunt and Minstrell (1994), and Smith (1990) have shown that prior misinformation detracts from learning new information and impedes complex reasoning.

Thus, if you are teaching 9th grade Biology and want your learners to think like biologists, first make sure they have basic knowledge about plants, animals, cell structure, etc. Your physics students will not be good thinkers if they lack information about forces, motion, and gravity. Your best thinkers in geometry will be those who know about angles, parallel lines, tangents and secants. And, your best third grade math problem-solvers will be those who know their math facts.

What Is a Domain-Specific Knowledge Base?

Learners know lot of things about history, geography, movies, baseball, inventors, religions, and so on. If you could look inside someone's head to find out what they know, just as you might browse through an encyclopedia, what might you see? One thing that would immediately draw your attention is that American history is located in one spot, European history in another. The sciences have their niche as well as literature. The American history spot is further organized into related domains like the Revolutionary War, Reconstruction period, or The Roaring Twenties.

Cognitive neuropsychologists are pretty convinced that these knowledge domains and their sub domains are really networks of nerve cells with connections running from one neuron to another (Kosslyn and Koenig, 1992). With estimates of 100 billion of these nerve cells in the brain's cortex each of which has hundreds to thousands of connections associated with it, that leaves a lot of room for information to be stored.

5-4
Now, if you could surgically remove one domain of knowledge, say geology, put it under an electron microscope that was connected to a computer system which let you see finer and finer neural networks at the touch of a key, what might you see? Let's say that you're looking at the geology domain of an eighth grade learner after six weeks of instruction on rocks and how they are formed? What you might see is something like Figure 5.1 (from Champagne et al, 1981).

Insert Figure 5.1 about here

Notice its hierarchical arrangement. This domain-specific knowledge base contains facts about rocks like terms and their definitions (metamorphic, igneous, sedimentary, granite). It contains concepts or instances of igneous rocks clustered together and separated from instances of metamorphic rock and its exemplars. Concepts are connected to form principles of how rocks are formed: weathering changes metamorphic rock into sedimentary rock.

What does this learner know about rocks? Quite a bit! First, as we pointed out, he knows the names of different rocks (limestone, sedimentary, shale, pumice, igneous). But not only does he know facts about rocks, he has concepts: metamorphic, igneous, and sedimentary are different kinds of rocks and there are examples of each kind. He also knows rules or principles about how one type of rock is formed by another and where different categories of rocks come from (igneous rock comes from volcanoes; sedimentary rock results from "stuff" that settles from air or water).

Every domain specific knowledge base has a similar structure: facts, concepts, and connections among concepts that lead to rules and
generalizations. In other words, a knowledge base contains every bit of information that we have ever learned about something and the connections among these bits. These knowledge bases are dynamic and ever changing. They increase in size as the student acquires more information. But more importantly they become more organized which allows the learner to learn new information and remember old information.

To give you an appreciation for this changing nature of knowledge networks, let's look at Figure 5.2 representing this same eighth graders rock knowledge base, but before the six week unit on rocks and how they are formed. What's different? What does this knowledge base tell us about what the student knows about rocks prior to instruction? He already knows somethings about rocks: that there are different kinds and that he can name some of them; but he also has some misinformation. He classifies lava as an example of sedimentary rock and granite as a metamorphic rock; he knows that slate is a rock but not it's category. He has little knowledge of the dynamics of rock formation. There are far fewer connections among the concepts indicating that the student is unaware of the relationships among the different rocks and how one type of rock is formed from another.

Insert Figure 5.2 about here

What can we conclude about the characteristics of a knowledge base from the above example? First, a knowledge base has breadth: it contains a wide array of information, concepts, rules, etc. Secondly, it becomes gradually more elaborated. It acquires more information and ideas. The learner associates more and more examples with particular concepts and
more and more concepts with one another. And third, it has a certain hierarchical organization that changes over time in the direction of dividing information into more subsets and establishing the relationship among the subsets.

The changes in the eighth grade students' rock knowledge base help us understand what it means to "know more." When we say that someone knows more about rocks after instruction than before, or that an expert race handicapper knows more than a novice, or that a college English major knows more than an eleventh grade American Literature student we mean that they have more bits of information. But, also, we mean that this information is more elaborated and organized, as in Figure 5.1. And, it is precisely this elaboration and organization that allows more knowledgeable learners to think and problem solve better in a given domain than the less informed learner regardless of IQ.

How Does a Knowledge Base Develop?

What causes domain-specific knowledge bases like the one we saw above to acquire more information and to become more elaborated and organized? Are these changes that you can bring about? Are you primarily in control of the learning or is the student? The answer is both but modern cognitive learning theory, as we saw in Chapter 3, stresses the role of the learners themselves in constructing knowledge networks.

Why does cognitive learning theory hold as fundamental the principle that learners and not teachers construct knowledge networks? For one, as Figure 5.2 shows, this eighth grade learner's knowledge of rocks before instruction already reveals breadth, elaboration, and organization. It got this way because the mind spontaneously constructs knowledge bases in this way.
The hierarchical organization is a natural transformation that occurs without the learner even being aware of it. The hierarchy wasn't created by systematically organizing lessons to present a fixed sequence of facts, discriminations, concepts, rules, and generalizations as behaviorists used to recommend. Learners create these organized networks on their own.

What is the role of instruction? It is to arrange learning environments so that learners get access to information and have the opportunities to relate it to what they already know. They will acquire pieces of information about rocks, associate these bits of information with the pieces they already know to form concepts and connections among concepts, and gradually develop deeper and deeper understanding with support and guidance from the teacher. Where initially their knowledge domains contained information that was organized based on surface features (e.g. the texture or color of rocks), eventually it becomes organized on the bases of more abstract principles (the way rocks are formed). And, it is the learner who constructs this organization.

Assessing What Learners Know

Given that learners construct their own knowledge networks or domains from classroom and other experiences, the assessment role of the classroom teacher is clear: monitor knowledge construction. This involves assessing prior information, the acquisition of new information (breadth), and its transformation (elaboration and organization). Using the rock knowledge example, let's say that the goal of the six-week geology unit was to develop deep understanding of rocks and how they are formed, and that the principle generalization to be learned is that the rocks which make up the earth's crust all have a common source: volcanic activity.
Students come to this unit knowing somethings about rocks but much of what they know are unconnected, isolated pieces of information some of which is misinformation. Many naive learners believe that all rocks are essentially the same. They look different because of the effects of weathering. Others believe that all rocks were formed under a vast ocean that covered the earth eons ago. Others, who know about volcanic activity, when asked why different rocks are heavier than one another, or have larger crystals than others, reason "well that's what they looked like before the volcanoes threw them out." The notion that the rate of cooling can cause molten rock to be glass or basalt is absent and, to them, counterintuitive.

These misconceptions can be a serious impediment to the types of learning and conceptual change that the teacher desires. In order to promote and monitor conceptual change, therefore, the teacher must assess not only his learner's new information about rocks and its changing elaboration and organization, but prior information and it's conceptual structure as well.

Thus, to assess what learners know, is to do the following:

1. Assess simple factual knowledge both prior to, during, and following instruction.

2. Assess elaboration as new pieces of information become added to what is already there to help form concepts. We'll call this assessing simple or surface understanding.

3. Assess organization as concepts become connected to one another to form rules, principles, and generalizations.

ASSESSING SIMPLE FACTUAL KNOWLEDGE

We use the expression "simple factual knowledge" to mean the facts, dates, names, expressions, or labels a person knows without much concern...
for depth of understanding or ability to use this information. Before you teach a unit on rocks, or fractions, or political parties, you want to determine whether they know such things as the names of different rocks, how to write a fraction, what a political party is and what nomination means. As your lesson progresses you give tests or ask questions to determine whether the knowledge base is growing, or is acquiring more breadth.

There are a number of issues that you should consider when assessing the extent of someone's factual knowledge (these issues also are germane to assessing simple understanding). Let's go over each of these and suggest ways of resolving them. Then, we will show you how to test for simple factual knowledge.

The issues you must resolve are: sampling, assessing recall versus recognition, reliability, construct relevancy, and efficiency.

Sampling

There are more bits of information that learners know before or after a lesson than you can practically assess. Biology contains more scientific terms, U.S. history more names and dates, Language Arts more spelling vocabulary, and basic math more number facts than could possibly be measured on some paper and pencil test. So, the question becomes: "Which information do I include on my test, and how many questions need I ask?

The issue pertains to the notion of sampling. Just as we sample voters to make inferences about how the nation feels about a particular political candidate, so also do we sample information to make inferences about what someone knows. In fact, any test is really just a sample of what someone knows. The trick is in drawing the best possible sample or asking the right questions to allow inferences about the knowledge base.
Sometimes knowledge domains are so finite that we don't need to sample. In geometry, for example, there are few enough types of angles, or geometric shapes that we can ask questions about all of them. There are several common misconceptions that learners have about plants and how they make food. You can assess all of these and changes in the misconceptions after a unit on photosynthesis. But, some knowledge domains are too vast to be exhaustively surveyed. Math facts is a good example, as is spelling, reading vocabulary, or grammar rules. In such cases we take a random sample of math facts or vocabulary and if a learner gets only 70% correct or 90% correct we infer a moderate or extensive knowledge base.

Thus, when knowledge bases are too extensive to ask all possible questions and we want to draw inferences about how informed or knowledgeable someone is, we have to sample from the knowledge domain and write questions about the sample. But the question remains, "Why is it important to make inferences about what someone knows?" This is the crux of the sampling issue. In some cases it is important, but in most cases it is not.

From a cognitive learning perspective, knowledge is not important for its own sake, i.e., knowing for the sake of knowing, unless you aspire to be an expert at Jeopardy or Trivial Pursuit. Knowledge is important because it is the foundation of deep understanding and complex problem solving. We want to assess a learner's knowledge base about rocks because knowing about certain terms, concepts and their connections is important for deep understanding and problem solving in this area.

The question you must ask in deciding which knowledge to assess is: What knowledge is essential to help the learner solve the important problems that are the focus of my unit on rocks, magnets, division, political parties,
fractions, plants, or poetry? When this knowledge is limited, as in our rock example, you need not sample. You can assess all the important pieces of information. But when this knowledge is extensive, as is the case of vocabulary for reading comprehension, math facts for automatically solving division problems, grammar rules for fluent writing, you must sample and make inferences about how much someone knows.

In Chapter 3 we discussed the significance of goals and objectives for teaching and testing. These goals and objectives will help you decide which knowledge is and is not essential to accomplish them. If your goal is deep understanding of the Depression, and you've defined the important knowledge as that which is contained in Long's and Roosevelt's speeches, then you can ask questions which comprehensively cover all that knowledge. On the other hand, if you want to measure learners' acquisition of knowledge from readings, class lectures, and textbook chapters, then you will have to ask a sample of all questions that could be asked.

Recall Versus Recognition

If you are trying to determine the extent of someone's spelling knowledge before or after a spelling unit, you can use either of these procedures:

1. Give the students a list of words and have them circle all those that are spelled incorrectly; or
2. Dictate the words and have the students spell them on a piece of paper.

Both techniques assess knowledge. The former measures recognition and the latter recall. The latter is a more demanding task for most learners.
than the former. Which is the better measure of knowledge? Before answering this question, here's another example.

How would you assess whether students know the names of famous explorers and their discoveries? Here are three ways:

(1) List the names of 10 explorers and their discoveries.
(2) Answer True or False: DeGama explored the Amazon River.
(3) The Spanish explorer who discovered the Mississippi River was named ________.

While all assess knowledge, (1) and (3) require recall, while (2) requires recognition. If you want to assess simple factual knowledge how do you know which one to use?

Tasks that ask learners to fill in blanks, list things in order, write down words, outline, recite facts, names, or dates require recall. Recognition is elicited by tasks that ask learners to do things like choose, select, or match. When is one task preferred over another?

The answer depends on your goal. In spelling, you may want learners to recognize that when something looks wrong after they have written it to use a dictionary. In this case, a recognition test will meet your needs. On the other hand, if you want your learners to write quickly, automatically, and correctly without depending on a dictionary, then the recall test is the better choice. In math if the learner will have access to calculators, then recalling number facts, or steps in solving a problem is less a concern than if your goal is to be able to provide quick and automatic answers to basic skill problems.

In general, when deciding whether to use a recall or recognition test to assess knowledge, remember that the former is more demanding than the latter. Then ask yourself whether, given the goals of the lesson, is it more
important that learners be able to recall information from memory, or recognize the presence or accuracy of something when they see it. Is it more important that they have the information at their finger tips or that they know where to get it when they need it? These are questions you must answer before you design your knowledge assessment.

Ways Of Measuring Recall And Recognition Of Simple Factual Knowledge

Simple factual knowledge is typically assessed with what are called "objective-type" item formats, with the word objective referring to the objectivity of scoring (i.e., avoiding subjectivity or bias). These formats include: true-false, completion, matching, and multiple-choice. Let's look at some rules for developing test items that measure simple factual knowledge using these formats.

**True-false Items.** True/false items are popular because they are quick and easy to write, or at least they seem to be. True/false items do take less time to write than good objective items of any other format, but good true/false items are not so easy to prepare.

As you know from your own experience, every true/false item, regardless of how well or poorly written, gives the student a 50% chance of guessing correctly even without reading the item! In other words, on a 50-item true/false test, we would expect individuals who were totally unfamiliar with the content being tested to answer about 25 items correctly. Fortunately, ways exist to reduce the effects of guessing. Here are some:

- Encourage *all* students to guess when they do not know the correct answer. Because it is virtually impossible to prevent certain students from guessing, encouraging all students to guess equalizes
the effects of guessing. The test scores will then reflect a more or less equal “guessing factor” plus the actual level of each student’s knowledge. This also will prevent test-wise students from having an unfair advantage over students who are not test-wise.

- Require revision of statements that are false. In this approach, provide space at the end of the item for students to alter false items to make them true. Usually the student is asked to first underline or circle the false part of the item and then add the correct wording, as in these examples:

  T  F  High IQ students always get good grades
        do not always

  T  F  Panama is north of Cuba
        south

  T  F  September has an extra day during leap year
        February

With this strategy, full credit is awarded only if the revision is correct. The disadvantage of such an approach is that more test time is required for the same number of items, and scoring time is increased.

Here are some suggestions to keep in mind when writing true/false test items:

- Tell students clearly how to mark true or false (for example, circle or underline the T or F) before they begin the test. Write this instruction at the top of the test also.
- Construct statements that are definitely true or definitely false, without qualifications. If the item is true or false on the basis of
someone’s opinion, identify the opinion’s source as part of the item—for example, “According to the head of the AFL-CIO, worker’s compensation is below desired standards.”

- Keep true and false statements at approximately the same length, and be sure that there are approximately equal numbers of true and false items.
- Avoid using double-negative statements. They take extra time to decipher and are difficult to interpret. For example, avoid statements such as “It is not true that addition cannot precede subtraction in algebraic operations.”
- Avoid terms denoting indefinite degree (for example, large, long time, regularly), or absolutes (never, only, always).
- Avoid placing items in a systematic pattern that some students might detect (for example, True-True-False-False, TFTF, and so on).
- Don’t take statements directly from the text without first making sure that you are not taking them out of context.

Completion Items. Like true/false, completion items are relatively easy to write. The first tests constructed by classroom teachers and taken by students often are completion tests. Like items of all other formats, there are good and poor completion items. Here are some suggestions for writing completion items:

- Require a single-word answer or a brief, definite statement. Avoid items so indefinite that they may be logically answered by several terms:

  Poor item: World War II ended in
Better item: World War II ended in the year

- Be sure the item poses a problem. A direct question often is better than an incomplete statement because it provides more structure for an answer.

Poor item: The main character in the story "Lilies of the Field" was called ____.

Better item: Who was the main character in the story "Lilies of the Field"?

- Be sure the answer is factually correct. Precisely word the question in relation to the concept or fact being tested. For example, can the answer be found in the text, workbook, or class notes taken by students?

- Omit only key words; don't eliminate so many elements that the sense of the content is impaired.

Poor item: The ____ type of test item usually is graded than the ____ type.

Better item: The multiple-choice type of test item usually is graded more objectively than the ____ type.

- Word the statement so the blank is near the end. This prevents awkward sentences.

- If the problem requires a numerical answer, indicate the units in which it is to be expressed (for example, pounds, ounces, minutes).

Matching Items. Like true/false, matching items are a popular and convenient testing format. Like good true/false items, however, good matching items are not easy to write. Imagine you are back in your ninth-grade American history class and the following item shows up on your test:
**Directions:** Match A and B

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Lincoln</td>
<td>a. President during the twentieth century</td>
</tr>
<tr>
<td>2. Nixon</td>
<td>b. Invented the telephone</td>
</tr>
<tr>
<td>3. Whitney</td>
<td>c. Delivered the Emancipation Proclamation</td>
</tr>
<tr>
<td>4. Ford</td>
<td>d. Only president to resign from office</td>
</tr>
<tr>
<td>5. Bell</td>
<td>e. Black civil-rights leader</td>
</tr>
<tr>
<td>6. King</td>
<td>f. Invented the cotton gin</td>
</tr>
<tr>
<td>7. Washington</td>
<td>g. Our first president</td>
</tr>
<tr>
<td>8. Roosevelt</td>
<td>h. Only president elected for more than two terms</td>
</tr>
</tbody>
</table>

See any problems? Compare the problems you identify with the descriptions of faults that follow.

**Homogeneity.** The lists are not homogeneous. Column A contains names of presidents, inventors, and a civil-rights leader. Unless they are specifically taught as a set of related men or ideas, this is too wide a variety for a matching exercise.

**Order of Lists.** The lists are reversed: column A should be in place of column B, and column B should be in place of column A. As the exercise is now written, the student reads a name and then has to read through all or many of the more lengthy descriptions to find the answer—a much more time-consuming process. It also is a good idea to introduce some sort of order—chronological, numerical, or alphabetical—to your list of options. This saves the student time.
*Easy Guessing.* Notice that there are equal numbers of options and descriptions. This increases the chances of guessing correctly through elimination. If there are at least three more options than descriptions, the chance of guessing correctly is reduced to one in four.

*Poor Directions.* The instructions are much too brief. Matching directions should specify the basis for matching: "Column A contains brief descriptions of historical events. Column B contains the names of U.S. presidents. Indicate who was president when the historical event took place by placing the appropriate letter to the left of the number in column A."

*Multiple Correct Responses.* The description “President during the twentieth century” has three defensible answers: Nixon, Ford, and Roosevelt. Also, always include first and last names to enhance recall and avoid ambiguities. Here is a corrected version of these matching items:

<table>
<thead>
<tr>
<th>Column A</th>
<th>Column B</th>
</tr>
</thead>
<tbody>
<tr>
<td>____1. Only president not elected to office</td>
<td>a. Gerald Ford</td>
</tr>
<tr>
<td>____2. Delivered the Emancipation Proclamation</td>
<td>b. Thomas Jefferson</td>
</tr>
<tr>
<td>____3. Only president to resign from office</td>
<td>c. Abraham Lincoln</td>
</tr>
<tr>
<td>____4. Only president elected for more than</td>
<td>d. Richard Nixon</td>
</tr>
<tr>
<td>two terms</td>
<td>e. Franklin Roosevelt</td>
</tr>
<tr>
<td></td>
<td>f. Theodore Roosevelt</td>
</tr>
<tr>
<td></td>
<td>g. George Washington</td>
</tr>
<tr>
<td></td>
<td>h. Woodrow Wilson</td>
</tr>
</tbody>
</table>
Notice that we now have complete directions, more options than descriptions, homogeneous lists (all items in Column A are about U.S. presidents and all the items in Column B are names of presidents), and we have made the alternatives unambiguous.

Here are some suggestions to keep in mind when writing matching items:

- Keep both the descriptions list and the options list short and homogeneous. They should fit together on the same page. Title the lists to ensure homogeneity (e.g., Column A, Column B) and arrange the options in a logical (e.g., alphabetical) order.
- Make sure that all the options are plausible “distracters” (wrong answer choices) for each description to ensure homogeneity of lists.
- The descriptions list should contain the longer phrases or statements, while the options should consist of short phrases, words, or symbols.
- Number each description (1, 2, 3, etc.) and letter each option (a, b, c, etc.).
- Include more options than descriptions, or some that match more than one, or both.
- In the directions, specify the basis for matching and whether options can be used more than once

**Multiple-Choice Items.** Another popular item format is the multiple-choice question. Multiple-choice tests are more common in high school and college than in elementary school. When writing multiple-choice items, be
careful not to give away answers by inadvertently providing students with clues in the following ways:

**Stem Clue.** The statement portion of a multiple-choice item is called the *stem*, and the answer choices are called *options* or *response alternatives*. A stem clue occurs when the same word or a close derivative occurs in both the stem and an option, thereby cluing the test taker to the correct answer. For example:

The free-floating structures within the cell that synthesize protein are called _____.

A. chromosomes  
B. lysosomes  
C. mitochondria  
D. free ribosomes

In this item the word *free* in the option is identical to *free* in the stem. Thus, the wise test taker has a good chance of answering the item correctly without mastery of the content being measured.

**Grammatical Clue.** Consider this item:

U.S. Grant was an ______.

A. army general  
B. navy admiral  
C. cavalry commander  
D. senator

5-21
Most students would pick up on the easy grammatical clue in the stem. The article *an* eliminates options a, b, and d, because "*an* navy admiral," "*an* cavalry commander," or "*an* senator" are ungrammatical. Option c is the only one that forms a grammatical sentence. A way to eliminate the grammatical clue is to replace *an* with *a/an*. Similar examples are *is/are, was/were, his/her*, and so on. Alternatively, place the article (or verb, or pronoun) in the options list:

Christopher Columbus came to America in ____.

A. a car  
B. a boat  
C. an airplane  
D. a balloon

**Redundant Words/Unequal Length.** Two very common faults in multiple-choice construction are illustrated in this item:

When 53 Americans were held hostage in Iran, ____.

A. the United States did nothing to free them  
B. the United States declared war on Iran  
C. the United States first attempted to free them by diplomatic means and later attempted a rescue  
D. the United States expelled all Iranian students

The phrase "the United States" is included in each option. To save space and time, add it to the stem: "When 53 Americans were held hostage in Iran, the United States ____." Second, the length of options could be a giveaway. Multiple-choice item writers have a tendency to include more information in the correct option than in the incorrect options. Test-wise students know
that the longer option is the correct one more often than not. Avoid making correct answers more than one and a half times the length of incorrect options.

_All of the Above/None of the Above._ In general, use “none of the above” sparingly. Some item writers use “none of the above” only when there is no clearly correct option presented. However, students catch on to this practice and guess that “none of the above” is the correct answer without knowledge of the content being measured. Also, at times it may be justified to use multiple correct answers, such as “both a and c” or “both b and c.” Again, use such options sparingly, because inconsistencies can easily exist among alternatives that logically eliminate some from consideration. Avoid using “all of the above,” because test items should encourage discrimination, not discourage it.

**APPLICATION**

In the spaces below select from the test item formats above two in which to demonstrate your skill at writing test items that measure recognition and another two that measure recall. Make each of your four test items measure simple factual knowledge in a subject you will be teaching.

Recall:
1.

5-23
Multiple-Choice Questions Beyond the Knowledge Level.

Multiple-choice items are unique among objective test items because if properly written they enable you to measure some limited types of higher-level cognitive objectives. Unfortunately, most multiple-choice items also are written at the knowledge level in the taxonomy of educational objectives. As a new item writer, you will tend to write items at this level, but you will need to write some multiple-choice items that measure cognitive objectives beyond the knowledge level. Following are some suggestions that can help you write multiple-choice questions beyond the knowledge level. Following this we will show you how to measure more complex learning with the essay format and in Chapter 8, we will show you how to measure deep understanding with performance-based assessments.

Use Justification to Assess Reasons Behind an Answer. Questions that follow a multiple-choice item can ask for specifics as to why a particular answer was chosen: For example:

Directions: Choose the most appropriate answer and cite evidence for your selection in the space below.

The principal value of a balanced diet is that it _____.

A. increases your intelligence
B. cures disease
C. promotes mental health
D. promotes physical health
E. improves self-discipline

What evidence from the text did you use to choose your answer?

Use Pictorial, Graphical, or Tabular Stimuli. Pictures, drawings, graphs, and tables can require the student to think at least at the application level in the taxonomy of educational objectives and may involve even higher cognitive processes. Also, such stimuli often can generate several higher-level multiple-choice items, as the questions below and Figure 5.3 illustrate:

Which of the following cities would be the best location for a steel mill?

A. Li (3A)
B. Um (3B)
C. Cot (3D)
D. Dube (4B)

Approximately how many miles is it from Dube to Rag?

A. 100 miles
B. 150 miles
C. 200 miles
D. 250 miles

In what direction would someone have to travel to get from Wog to Um?
A. northwest
B. northeast
C. southwest
D. southeast

Insert Figure 5.3 about here

Use Analogies to Show Relationships Between Terms. To answer analogies correctly, students must not only be familiar with the terms but be able to understand how the terms relate to each other. For example:

Physician is to humans as veterinarian is to
A. fruits
B. animals
C. minerals
D. vegetables

Require Application of Principles or Procedures. To test whether students comprehend the implications of a procedure, have them use the principle or procedure with new information or in a novel way. This requires them to do more than just follow the steps in solving a problem. It has them demonstrate an ability to go beyond the context within which they originally learned a principle or procedure. Consider this example from a division lesson on computation of ratios and proportions:
After filling the car’s tank with 18 gallons of gasoline, Mr. Watts said to his son, “We’ve come 450 miles since the last fill-up. What kind of gas mileage are we getting?” Which is the best answer?

A. 4 miles per gallon
B. 25 miles per gallon
C. Between 30 and 35 miles per gallon
D. It can’t be determined from the information given

This item tests not only knowledge of division but also application skills.

Here are some suggestions to keep in mind when writing multiple-choice items:

• Be sure that there is one and only one correct or clearly best answer.

• Be sure all wrong answer choices (“distracters”) are plausible. Eliminate unintentional grammatical clues, and keep the length and form of all the answer choices equal. Rotate the position of the correct answer from item to item randomly.

• Use negative questions or statements only if the knowledge being tested requires it. In most cases it is more important for the student to know what the correct answer is rather than what it is not.

• Include three to five options (two to four distracters plus one correct answer) to optimize testing for knowledge, comprehension, or application rather than encouraging guessing. It is not necessary to provide additional distracters for an item simply to maintain the same number of distracters for each item.

• Use the option “none of the above” sparingly and only when all the answers can be classified unequivocally as wrong.
• Avoid using “all of the above.” It usually is the correct answer and makes the item too easy for students who have only partial information.

APPLICATION

In the spaces below write two multiple choice items that demonstrate your ability to use a multiple-choice format to achieve some types of higher order thinking.

1.

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

2.

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

5-29
Reliability

While answers to the above objective test items are easier to score than answers to essay questions, they still present significant problems from the standpoint of reliability. The reliability problems stem from the poor way in which objective items are often written. When such questions are poorly written, conclusions about what someone does or does not know will not be valid. Here are some of the more common mistakes that you will want to avoid:

1. **Ambiguous or unclear presentation of question.** Write test instructions and questions in simple, uncomplicated language. This way, learners won't have to struggle to untangle complicated phrases or clauses. When in doubt have a student in a lower grade read the item and tell you what the question means.

2. **Not enough questions.** Include enough questions on the test to adequately cover all that the learner is expected to know. The more items, the greater the sample of performance obtained and the higher the test reliability.

3. **Not enough time.** Allow students sufficient time to take the test. A rushed student is more likely to make foolish errors, which will lower the reliability of the test.

4. **Uncomfortable conditions.** Make sure the testing conditions (temperature of the room, noise level, seating arrangements) are conducive to maximum performance. If the test situation is uncomfortable, learners are more easily distracted and less likely to demonstrate what they really know.
(5) **No test blueprint.** Follow a test blueprint so that you adequately sample all that has been taught--and all that the learner knows. The test blueprint will enable you to construct enough test items to adequately measure what you've taught.

(6) **Ambiguous answers.** Write objective questions that have easily identifiable right and wrong answers. When using restricted-response questions, prepare a model answer or scoring guide before grading. This way the correct answer will be scored "right" every time, or a "good" answer given a high rating, because it reflects your scoring guide, not your mood or temperament at the time of grading.

There are other common kinds of mistakes that lower the reliability of tests that measure simple knowledge. For each item type there are simple rules to follow when writing to avoid these mistakes. These rules and methods that you can use to determine the reliability of your objective type test items are contained in the suggested readings at the end of this chapter.

**Construct Relevancy**

The principal threat to construct relevancy for objective type items is the reading level of the question. If the reading level is too high, then your test will be more a measure of reading comprehension than knowledge of rocks, for example. Often children cannot demonstrate their knowledge of math because they don't understand the question that they are reading. Unless you specifically intend to measure reading ability, then the reading level of the test should be below that of the learners who are taking the test.

A test is valid when it measures what is purports to measure. Construct relevancy is one method of determining a test's validity. Other
methods that can help you determine the validity of your objective type test items are contained in the suggested readings at the end of this chapter.

Efficiency

Efficiency refers to ease of writing, ease of scoring, and content coverage. The easiest items to write among those measuring simple knowledge are the fill-in-the-blank variety. Conversely they are hardest to score since we can't always anticipate the range of plausible words that students put in blanks. Multiple-choice items, on the other hand, are among the most difficult items to write but among the easiest to score. It takes longer to read and respond to a multiple-choice question than it does to a true-false item. Consequently you get more coverage of the knowledge domain with the latter item type in a 20 minute testing period than you do with the former.

The issues confronting you when you assess simple factual knowledge often represent tradeoffs. While some item formats may give you greater content coverage (like true-false), they are also more susceptible to guessing. While multiple-choice items are less susceptible to guessing, they take longer to write. Completion items require recall and are less susceptible to guessing, but they are more difficult and time consuming to score. Table 5.1 lists the advantages and disadvantages of each objective-type item vis-a-vis some of the issues raised above.

Insert Table 5.1 about here

ASSESSING SIMPLE UNDERSTANDING

5-32
When we measure simple understanding we want to determine more than whether a learner knows the names of different kinds of rocks, the parts of a leaf, the names of atomic elements, or can give a definition of a vocabulary word. We want to ascertain if they have acquired a concept: whether they know what the word or expression means. In other words, can they give examples of it, tell how it is different from or similar to another term or expression, explain what it means using their own words, etc.? What we are doing is determining if they know or represent information as an idea or simply know it by rote memorization.

We use the expression "simple understanding" to make the point that there are different degrees of understanding--that it deepens as we learn more and increase the complexity of the knowledge base. For example, your understanding of the expression "authentic learning assessment" is richer and deeper now than when you first began to read this book. You associate it with more terms like measurement, evaluation, test, reliability, validity, etc. You link it with a concept called motivation. You are ready to begin to build an authentic learning assessment system. In other words your understanding is becoming deeper. So, too, a learner's understanding of the word "rock" deepens as she learns more about rocks. We will discuss ways to assess deep understanding in subsequent chapters.

You can further appreciate the distinction between simple and deeper understanding by examining the eighth grader's pre-instruction rock knowledge shown in Figure 5.2. He knows the terms lava and granite. He knows that they are the names of different kinds of rocks. But his concept or understanding of these rocks is incomplete. He classifies granite as metamorphic when it is igneous. He classifies lava as sedimentary when it also is igneous. He is not sure what slate is. It appears as if he doesn't really
understand any of these terms as concepts—the true distinguishing features of the rocks and how they are similar and different. But after instruction (Figure 5.1), we see all rocks correctly classified indicating that he understands the terms and their relationships to one another.

You can assess simple understanding of terms, expressions, etc. using higher order multiple choice questions or restricted response essay questions. Restricted response essay questions pose a simple problem to a student. In order to answer the item, the student must recall information and explain something in their own words within the specific limits posed by the test question (Kubiszyn & Borich, 1996). For example, the following is a restricted response question assessing whether a learner understands the concepts "igneous" and "metamorphic":

In your own words, explain two differences between igneous and metamorphic rock and give one example of each not given in class.

In order to demonstrate that they know igneous and metamorphic as concepts and not just simple knowledge, learners must do two things: they must use their own words to explain the differences and not simply recall what their text said or what they copied from an overhead; and they must give original examples of each rock. If they can do this, then the teacher can correctly judge that they have the concepts. In this section, we will explain what restricted-response essay items are and provide suggestions for writing them.

The Restricted-Response Essay
An essay that poses a specific problem for which the student must recall proper information, organize it in a suitable manner, derive a defensible conclusion, and express it according to specific criteria is called a restricted-response essay item. The statement of the problem specifies response limitations that guide the student in responding and provides evaluation criteria for scoring. For example:

List the major similarities and differences between U.S. participation in the Korean War and World War II, being sure to consider political, military, economic, and social factors. Limit your answer to one page. Your score will depend on accuracy, organization, and conciseness.

An essay that allows the student to determine the length and complexity of a response is called an extended-response essay. This type of essay is most useful at the analysis, synthesis, and evaluation levels of cognitive complexity and for assessing deep understanding. We will have more to say about extended-response essays in our chapters on performance assessment and portfolios.

When Should Restricted Essay Questions be Used?

Some of the conditions for which restricted essay questions are best suited are when:

- The instructional objectives require supplying information rather than simply recognizing information. These processes often cannot be measured with objective items.
• Relatively few areas of content need to be tested. If you have 30 students and design a test with six restricted-response essays, you will spend a great deal of time scoring. Use restricted essays when class size is small, or use them in conjunction with objective items.

• Test security is a consideration. If you are afraid test items will be passed on to future students, it is better to use an essay test format. In general, a good essay test takes less time to construct than a good objective test.

Here are some learning outcomes for which restricted essay items may be used:

- Analyze relationships
- Compare positions
- State necessary assumptions
- Identify appropriate conclusions
- Explain cause-and-effect relations
- Formulate hypotheses
- Organize data to support a viewpoint
- Point out strengths and weaknesses
- Integrate data from several sources
- Evaluate the quality or worth of an item, product, or action

Here are some suggestions to keep in mind for writing restricted-response essay items:

• Have clearly in mind what mental processes you want the student to use before starting to write the question. Refer to the mental processes required at the various levels in the taxonomy of
educational objectives for the cognitive domain (for example, compare and contrast, organize, demonstrate, develop). For example, if you want students to apply what they have learned, determine what mental processes would be needed in the application process.

*Poor item:* Criticize the following speech by our President.

*Better item:* Consider the following presidential speech. Focus on the section dealing with economic policy and discriminate between factual statements and opinion. List these statements separately, label them, and indicate whether each statement is or is not consistent with the President’s overall economic policy.

- Write the question to clearly and unambiguously define the task to the student. Tasks should be explained (1) in the overall instructions preceding the test items and/or (2) in the test items themselves. Include instructions for the type of writing style desired (for example, scientific vs. prose), whether spelling and grammar will be counted, and whether organization of the response will be an important scoring element. Also, indicate the level of detail and supporting data required.

*Poor item:* Discuss the value of behavioral objectives.

*Better item:* Behavioral objectives have enjoyed increased popularity in education over the years. In your text and in class the advantages and disadvantages of behavioral objectives have been discussed. Take a position for or against the use of behavioral objectives in education and support your position with at least three of the arguments covered in class or in the text.
• Start restricted-response essay questions with such words or phrases as compare, contrast, give reasons for, give original examples of, predict what would happen if, and so on. Do not begin with such words as what, who, when, and list, because these words generally lead to tasks that require only recall of information.

  Poor item: List three reasons behind America’s withdrawal from Vietnam.

  Better item: After more than 10 years of involvement, the United States withdrew from Vietnam in 1975. Discuss what might have happened if America had not withdrawn at that time and had not increased significantly its military presence above 1972 levels.

• A question dealing with a controversial issue should ask for, and be evaluated in terms of, the presentation of evidence for a position rather than the position taken. It is not defensible to demand that a student accept a specific conclusion or solution, but it is reasonable to appraise how well he or she has learned to use the evidence upon which a specific conclusion is based.

  Poor item: What laws should Congress pass to improve the medical care of all citizens in the United States?

  Better item: Some feel that the cost of all medical care should be borne by the federal government. Do you agree or disagree? Support your position with at least three logical arguments.

• Establish reasonable time and/or page limits for each essay item to help the student complete the entire test and to indicate the level of
detail you have in mind. Indicate such limits either in the statement of the problem or close to the number of the question.

- Restrict the use of essays to learning outcomes that cannot be satisfactorily measured by objective items.

Scoring Restricted Response Essays

As mentioned, essays are difficult to score consistently across individuals. That is, the same essay answer may be given an A by one scorer and a B or C by another scorer. Or the same answer may be graded A on one occasion, but B or C on another occasion by the same scorer! As disturbing and surprising as this may seem, these conclusions were supported by research findings (Coffman, 1972). What can you do to avoid such scoring problems?

Write Good Essay Items. Poorly written questions are one source of scorer inconsistency. Questions that do not specify response length are another. In general, long (for example, three-page) essay responses are more difficult to score consistently than shorter responses (say, one page). This is due to student fatigue and subsequent clerical errors as well as to a tendency for grading criteria to vary from response to response, or for that matter, from page to page, or even paragraph to paragraph within the same response.

Use Several Restricted-Response Items. Rather than a single restricted-response essay, use several smaller essays. This will provide students a greater opportunity to show off their skills and a greater variety of criteria to respond to.
Use a Predetermined Scoring Scheme. All too often, essays are graded without the scorer having specified in advance what he or she is looking for in a “good” answer. If you do not specify the criteria beforehand, your scoring consistency will be greatly reduced. If these criteria are not readily available (written down) for scoring each question, the criteria themselves may change (you may grade harder or easier after scoring several papers, even if the answers do not change). Or your ability to keep these criteria in mind will be influenced by fatigue, distractions, frame of mind, and so on. Because we all are human, we all are subject to these factors.

APPLICATION

In the space below, write a restricted response essay to assess simple understand of a concept you will be teaching. Identify what behaviors you hope to assess with your item.

Restricted-response essay item:

_________________________________________________________________  
_________________________________________________________________  
_________________________________________________________________  
_________________________________________________________________  
_________________________________________________________________  

Behaviors you hope to measure with your restricted-response essay item:

5-40
ASSESSING KNOWLEDGE ORGANIZATION

As Figure 5.1 showed, the mind spontaneously organizes information as it is learned. As students attend and listen to lectures and discussions, or read from their textbooks, they link this new information with prior learning and this linking helps them to learn concepts, principles, and generalizations. Over time, as their knowledge base grows it becomes increasingly organized in a hierarchical fashion as shown in Figure 5.1.

Even though learners construct this organization on their own, there are a variety of things that teachers can do to facilitate it. At the start of a lesson, they can ask questions that get learners to recall previous learning. During the lesson teachers can ask questions and provide activities that help learners see similarities and differences and to detect patterns and relationships among the pieces of information that they are hearing or reading. Teachers can construct outlines or other schematics that visually remind them of how new information is organized and relates to previously learned information. Figures 5.4 and 5.5 represent visual plans constructed by teachers for two
interdisciplinary thematic units. These became important tools that help them organize knowledge for instruction in ways that emphasize the interrelationships that build to larger themes and concepts.

Insert Figures 5.4 & 5.5 about here

Thus, knowledge organization is an important goal of instruction because an organized knowledge base helps students acquire new information, learn it in a meaningful way, remember it, and better solve problems that require it. Assuming that knowledge organization is a goal for your learners, what are some ways to assess it?

First of all, assessing knowledge organization and concepts are not the same. The assessment procedures discussed above let you determine simple understanding of terms and expressions like photosynthesis, igneous, inclined plane, Romanticism, oligarchy, the Depression, etc. But they don't tell you much about how well the student understands the relationships among the concepts. Look again at the eighth graders rock knowledge before and after instruction. Before instruction, much of the knowledge was in pieces, but after instruction not only are the pieces there (the concepts), but these pieces have connections. It is these connections that we want to assess when we assess knowledge organization.

When we talk about connections between and among concepts, we are talking about the students knowledge and understanding of rules, principles, and generalizations. For example, a principle of rock formation is that sedimentary derives from igneous rock that has been affected by "stuff that settles from air or water." Or that metamorphic and sedimentary rocks are derived from igneous rocks as a result of weathering.
Learners construct these principles or generalizations as a result of instruction that allows them to gather information about rocks through direct experience, to explore similarities and differences, and to establish relationships and connections. Learners of all ages spontaneously organize information and form orderly knowledge bases in history, geography, literature, mathematics, etc.

Thus, when assessing for knowledge organization, what you are identifying is the connections among concepts, or the sets and subsets of knowledge. But how can learners display their organization of knowledge—of cause/effect relationships, similarity/contrasts, problem/solutions, etc.? The inability of traditional outlines to reveal knowledge organization, where major topics and subtopics are grouped in a I, II, III, A, B, C order, may explain why many learners discard them except when teachers require them. Some educators believe such outlines emphasize form over meaning (Goetz et al., 1992), since they can easily represent an imposed structure that differs from the way the knowledge should actually be organized for deep understanding.

Dansereau (1988) urges teachers to model alternate outline strategies for their learners in order to help them when they listen to lectures or read from books, which, in turn, can be used to reveal the depth of understanding and organization of their knowledge base. He advocates graphic outlines displayed as webs, much like the visual interdisciplinary thematic unit plans shown in Figures 5.4 and 5.5. Webbing is a free-form outline technique which learners can be asked to provide to display their level of understanding of lecture or textbook content, as displayed in Figures 5.6 a-c. Some rules for communicating to your learners how to construct webs, networks or maps are:

5-43
• Display only essential information or big ideas or concepts
• Assign the central idea or concept to a central location
• Draw arrows from the main ideas to show relationships
• Label the arrows with key words or code letters to describe the relationship.

Insert Figure 5.6 a-c about here

Summary

This chapter presented some of the ways you can assess your learners' simple factual knowledge and understanding. It presented formats for testing simple factual knowledge with the true-false, completion, matching and multiple choice item formats. It also presented formats for measuring simple understanding with the higher order multiple choice and restricted response essay format. Each of these assessment formats were examined for their sampling requirements, contribution to measuring recall versus recognition, their reliability, construct relevancy and efficiency. This chapter also presented several ways of measuring knowledge organization through the vehicles of visual webs, networks and maps that can capture the degree to which you have taught and your learners have acquired the relationships among bits of knowledge necessary for building toward higher order concepts and themes.

Activities

5-44
1. What implications for your classroom teaching and assessment can be drawn from Ceci's study of horserace handicapping with respect to the relative emphasis you might want to place on building your learners' domain-specific and domain-general knowledge base?

2. For your grade or subject, construct two items each that measure simple factual knowledge using the true-false, completion, matching, and multiple-choice formats. Critique each of your items using the criteria provided.

3. Select a topic for which you would like your learners to exhibit simple factual understanding. Using a higher-order multiple choice format, prepare two items that go beyond simple recall and recognition to elicit evidence of more complex learning.

4. Prepare two restricted response essay questions over a topic you will teach. Indicate the learning outcomes your essay items are intended to achieve and explain what you would do to assure that you will score your essays consistently across learners.

5. Describe an interdisciplinary thematic lesson you would like to teach that would require your learners to organize knowledge, i.e., understand relationships, make connections, understand causes and effects, etc. Then, from your learners' perspective prepare a web diagram that would exhibit to your satisfaction that they have organized the knowledge successfully.

Suggested Readings

Chapter 11 covers how you can assure the quality of your objective type questions by determining whether they are of appropriate difficulty and whether each item reflects the overall behavior that the test is presumed to measure. Chapter 12 covers rules for improving the reliability and validity of your objective test items.


Chapters 15 and 16 cover all of the methods traditionally used to determine the validity and reliability of objective tests and how to interpret their meaning.
Chapter 6
Assessing What Learners Can Do

In Chapter 5, you learned about methods for assessing what learners know. As you will recall, learners know facts, concepts, rules and generalizations. Their knowledge base is organized in a manner which makes it easier to remember and recall information. You can assess this knowledge base and its degree of organization using a variety of measurement techniques like completion, short-answer, true-false, multiple-choice items and concept maps.

Often teachers are interested in assessing not only what learners know, but also what they can do. They want to know if their learners can use the information which they possess. For example, it is important to assess what a learner knows about a microscope and whether he or she can use one. Learners should not only know about fractions but also how to add, subtract, and reduce them. They should know about the parts of speech and be able to write a sentence, paragraph, or essay.

It is possible to know about things and not be able to apply what you know. For example, you might be able to tell a friend how to bake bread yet be unable to do it yourself. Or, you might be able to describe clearly how to make parallel turns when skiing but fall head over-heels the first time you try yourself. Similarly, a learner may correctly state the rule for how to add fractions with unlike denominators but not be able to arrive at the right answer when applying the rule.

Cognitive psychologists, therefore, make a distinction between knowledge about things and knowledge of how to perform activities. As you may recall, the former is called declarative knowledge; the latter is called
procedural knowledge. The distinction is important because different teaching conditions are required for students to learn procedural knowledge in comparison to declarative knowledge. Different methods also are used to assess these knowledge bases. These methods are the focus of this chapter.

Procedural Knowledge

Karl Malone, one of the best professional basketball players of the past quarter century, is called the "Mailman," because he delivers. Experts in any field "deliver" just like Karl Malone. They not only know, but they also get the job done. To use a popular expression, they "talk the talk and walk the walk." Experts deliver because they have extensive, organized networks of procedural knowledge. They not only know "about" reading, writing, geometry, anatomy, and mathematics, but they also know "how" to read, produce a short story, bisect a line, dissect a specimen, and solve a quadratic equation.

Their knowledge of how to do things, of how to get the job done, is stored in long-term memory. In Chapter 5 we pointed out that declarative knowledge is also stored in long-term memory. Cognitive psychologists speculate that there are separate information stores or networks of neurons for declarative and procedural knowledge. They also speculate that the organizational properties of the two types of knowledge are different.

Declarative knowledge, as we have seen, appears to be organized hierarchically. Procedural knowledge, on the other hand, seems to be organized into networks of "if-then" or condition-action rules. The "if" component of the rule specifies the conditions that are necessary for a particular action to take place (the "then") component. For example, procedural knowledge about adding fractions, is organized into a network of
"if-then" rules: If the denominators are the same, then add the numerators; if the denominators are different, then first reduce the denominators to their least common multiple, etc.

Much of our procedural knowledge is used so automatically that we don't have to think consciously about it. Reading is an example. When students first learn to read, they consciously act out "if-then" rules: if the letters "t" and "h" are next to one another, then they make a "th" sound; if the vowels "e" and "a" are next to each other, they pronounce the combination as long "e," etc. But with practice, vowel and consonant combinations are read automatically. Cognitive psychologists speculate that procedural knowledge started as declarative knowledge (Anderson, 1983). In other words, we first learn about something. We learn about vowels and vowel sounds and combinations. Then we learn how to read them. We learn about parts of speech. Then we learn how to write a sentence. We learn about bisecting angels. Then we learn how to do it. We learn about swinging a golf club, or hitting a backhand in tennis, or doing a hand stand. Then we learn how to do these actions.

Although declarative knowledge and procedural knowledge occupy different locations in the brain's long-term memory stores and are organized in different ways, they work together to produce expert performance. Your classroom management skills depend on what you know about appropriate and inappropriate learner behavior. But, they also depend on how well you communicate rules, encourage good behavior, and prevent inappropriate behavior. Likewise you need extensive declarative knowledge about U.S. history to develop a lesson plan. But, you also must know how to get a learner's attention, model good thinking, coach learning, and fade out your direct involvement to allow for learner independence.
How Procedural Knowledge Develops

The development of procedural knowledge is one of the most important missions of schooling. Children should leave the early elementary grades able to read and write fluently—or automatically. They should be able to perform basic skill procedures of addition, subtraction, division, and multiplication without thinking. They should know how to read a story and comprehend what it is about. They should know how to use some basic strategies for solving math and science problems. When they reach middle and high school, they should learn procedures to automatically write extended essays, carry out experiments, analyze a piece of writing, etc.

Learners pass through a series of stages on their way to automatically carrying out procedural knowledge. These stages are called the cognitive stage, associative stage, and the autonomous stage (Gagne, Yekovich, and Yekovich, 1993).

Cognitive stage. During the cognitive stage, the learner is a novice and is dependent on declarative knowledge to cue her about how to do something. When she sees on a worksheet the problem:

\[ \frac{5}{6} + \frac{1}{3} = \]

she might have to consult a procedure that the teacher has given to her on a handout or is written in her textbook which says:

1. Find the least common denominator.
2. Divide the denominator of the second fraction into the
   least common denominator.
3. Multiply the result of step 2 by the numerator of the first
   fraction, etc.
The learner might then verbally state to herself each step before carrying it out. She consciously executes all the steps in the procedure.

If you have ever skied in the back country and are just learning to execute a turn, you probably start by saying to yourself:

*Now slide across the slope. Keep your weight on your downhill ski. Pick up and step out your uphill ski into a wedge. Point that ski in the direction you want to turn. Now plant your inside pole and shift all your weight onto your outside ski. Now pick up the inside ski, lay it in parallel to the outside ski and skid around the turn.*

*Keep your hands forward and knees flexed.*

At this point in your development, you move slowly, are conscious about things like weight shifts, hand positions, and pole placements and must cue yourself about every movement.

**Associative stage.** During this stage, the learner's performance gradually changes from a sequence of actions controlled by cues or reminders that are guided by her declarative knowledge to a sequence of actions that are less conscious. Performance slowly becomes smooth and rapid. Two changes bring this about. First, several steps are collapsed into one step. The learner combines several individual actions or movements. In the fraction example this could involve one set of cues that elicit steps 1-3 quickly, followed by another set of cues that elicit steps 4-6, etc., rather than a separate set of cues for each step. In the skiing example, it might involve one fluid movement to the point where the inside pole is planted.
The second change that occurs during this stage is the dropping out of the declarative knowledge cues. The learner no longer has to sit and think about what comes next, or hesitate for a second before executing the next movement. The learner no longer consciously searches for the next thing to do. Rather she automatically executes the action sequence, although not as smoothly as an expert.

**Autonomous stage.** This is a continuation of the associative stage. The action sequence is fine tuned. There is little hesitation. One action follows the other in a smooth, effortless, movement or motion. It is as if the learner is on automatic pilot. You can detect almost no thinking by the learner about what to do next. The only conscious thought is that of the goal: solving the problem, getting down the hill. Actions are so automatic that the learner can think about something else as she is performing them. While solving fractions, the learner thinks about recess as soon as the worksheet is done. The skier admires the breathtaking landscape as she plunges down the slope.

Types of Procedural Knowledge

You have learned that there are different types of declarative knowledge: facts, concepts, rules, generalizations. There are also different types of procedural knowledge that schools seek to develop. We will refer to these as (1) domain-specific basic skills; (2) domain-specific strategies; and (3) domain-general strategies (Gagne, Yekovich, and Yekovich, 1993).

**Domain-specific basic skills.** We introduced the concept of a knowledge domain in Chapter 5. As you will recall, a knowledge domain is
any defined subject-matter area that can vary in breath. Biology is a knowledge domain that is very broad. Zoology is a branch of biology that is narrower than biology but still considered a domain. American history is a knowledge domain that has many sub domains as does mathematics, or geography. Procedural knowledge that is specific to any of these domains or sub domains is referred to as domain-specific basic skills.

We have already given several examples of such procedural knowledge: adding fractions, bisecting an angle, making parallel skiing turns, baking bread. Other examples include three digit multiplication, computing a square root, proving a geometric theorem, constructing a map, focusing a telescope, tying shoelaces, riding a bicycle, starting a car, saving information to a floppy disc, taping a TV program on a VCR, etc.

As you can see, the goal of teaching domain-specific basic skills for learners is that they reach the autonomous stage of procedural knowledge development. In other words, once a learner recognizes that the task before him requires a certain sequence of actions, then he executes these actions inevitably and automatically. For example, once the learner recognizes the problem as one involving subtraction (the "if" part of procedural knowledge), he immediately compares the top and bottom numbers of the farthest right-hand column and either subtracts or executes a borrowing routine (the "then" or action component of procedural knowledge).

**Domain specific strategies.** When you teach domain-specific basic skills, your goal is smooth, efficient, automatic execution. But there are some procedures which you teach in a domain for which your goal is conscious, controlled, deliberate execution by the learner. These are called domain-specific strategies. For example, as part of writing instruction,
students learn to execute a routine to get them started that involves (1) generating a topic sentence, (2) noting reasons, (3) examining the reasons and asking if readers will accept each reason; (4) coming up with an ending (Scardamalia and Bereiter, 1986). This routine or strategy is domain specific because it is taught as a writing tool. But the learner should use it in a deliberate, consciously controlled manner.

Another example of a domain-specific strategy is safe driving techniques. As you drive you should consciously monitor traffic conditions like speed, the distance between your car and the ones in front of you, road conditions, congestion, etc. This strategy should be consciously and deliberately applied. Other examples of domain specific strategies are routines for setting up a laboratory experiment, solving word problems in mathematics, monitoring reading comprehension, developing a plan for a writing project, checking to see that subtraction problems have been computed correctly.

**Domain-general strategies.** The above examples of procedural knowledge are specific to areas of learning. There is another type of procedural knowledge that encompasses routines or sequences of actions that are useful in any domain. These are called domain-general strategies. For example, teachers always have as a goal the improvement of someone's ability to learn or think regardless of whether the learner is studying history, science, geography, or mathematics. The student should be able to set goals for learning, come up with a plan to achieve the goal, monitor learning, and evaluate her progress. What we have just described is a "learning to learn" strategy that a student can use regardless of the content area she is trying to master.
Robert Sternberg of Yale University and Howard Gardner of Harvard University have developed a program called "PIFS" (Practical Intelligence For Schools), the goal of which is to help students improve their learning in any school subject (Gardner et al, 1994). Other examples of programs to teach domain-general strategies are CoRT (Cognitive Research Trust), the goal of which is to teach general thinking skills to learners as young as 8 years old (Nickerson, 1988); and the Productive Thinking Program developed by Martin Covington of the University of California at Berkeley (Covington, Crutchfield, Davies, & Olton, 1974).

These programs attempt to teach learners a consciously controlled sequence of actions that they use whenever faced with any type of problem solving task. For example, when faced with any problem that requires a solution learners should:

1. First gather the facts
2. State the problem clearly
3. Brainstorm solutions
4. Think of unusual ways to solve it
5. Be systematic in solving the problem
6. When stuck, think of a new approach to the problem
7. Self-monitor progress and solutions

Domain-specific strategies and domain-general strategies are examples of what in Chapter 3 were called cognitive strategies. As we pointed out, they are important goals for any learning assessment program. They require different teaching conditions and assessment techniques as compared to domain-specific basic skills. Chapter 9 will be devoted to the assessment of these types of procedural knowledge. In the remainder of this chapter, we
will concern ourselves with procedural knowledge pertaining to automated basic skills.

Teaching Domain-Specific Basic Skills

We stated above that making distinctions among the different types of knowledge (declarative versus procedural; types of procedural knowledge) are more than just an academic exercise. The importance of these distinctions for day-to-day lesson planning is that each type of knowledge needs to be taught and assessed differently. Procedural knowledge in particular is challenging to teach because it involves not only teaching rules and sequences of actions, but also when and how to use these rules and actions automatically to solve problems. Learners must be able to identify the type of situation or problem, choose the appropriate sequence of actions to solve the problem, and execute the actions efficiently and effectively.

When teaching this type of procedural knowledge, an important thing to remember is the notion of knowledge construction. You should see your task more in terms of setting up meaningful situations where learners recognize the need for constructing a particular procedure or routine, as opposed to your training a rote sequence of actions. Learners must not only know how to perform a procedure, but also why. Motivation is a key factor in the learning and expert execution of procedural knowledge.

Once you have provided learners with a meaningful context in which to learn a specific procedure, then the challenge is to provide the conditions that allow them to attain the third or autonomous stage of procedural knowledge development. You will have to model the sequence of actions in a systematic, step-by-step manner. Learners must then imitate your actions and practice the steps under your guidance and coaching. Remember that the
purpose of practice is efficient and automatic use of the procedure. This will require extensive repetition in conjunction with your specific feedback. The more examples and contexts you provide, the more learners will recognize when and how to use the procedure.

Now let's turn to the assessment techniques which you should develop to help learners construct autonomous procedural knowledge.

Assessing Domain-Specific Basic Skills

While evidence of declarative knowledge can come in a variety of ways (short answer questions, objective-type test items, recall or recognition exercises, application tasks, etc.), procedural knowledge has to be performed. In other words it has to be observed in order to be assessed. The only way to know if someone can ride a bike is to see them do it. Their telling you how it's done isn't enough. You say you know how to manage classroom behavior? Well, let's see you do it? You say you can teach learners how to subtract? Well, show me!

We want to stress, however, that procedural knowledge, as we pointed out above, has a declarative knowledge foundation. Knowing how to subtract depends on knowing what subtraction is. Knowing how to manage classroom behavior depends on knowing something about behavior. Nevertheless, in this chapter, we will focus on the "how" aspect of procedural knowledge—on the ways to determine if learners efficiently and effectively know how to do something. In the next chapter, we will discuss ways to assess both the how and what of student knowledge using the techniques of performance assessment.

Assessing Processes and Products
A process is the sequence of actions, routines or steps that a learner takes to complete an activity. The product is the concrete, tangible outcome that results from those actions. In math, for example, the product might be the written solution to a geometric proof, with the process being the logical sequence of steps that the student followed to prove the theorem. A loaf of bread is a product. The process involves mixing the ingredients, kneading the dough, etc.

Most of the time teachers are concerned with assessing products: a book report, map, drawing, science project, or essay rather than the act of developing these concrete outcomes. It is usually easier to evaluate a product. You have a single object which you can compare with other products or to some standard. You can take your time while doing so. Processes, however, consist of a series of steps that you usually have to observe being performed.

Most of classroom assessment concerns itself with products. When the product is competently completed in a timely fashion, the correct processes were probably used. But, when the product is faulty, contains errors or mistakes, and took an inordinate amount of time to complete, a product assessment by itself does not tell why. This is where process assessment becomes important. It helps you diagnose reasons why the product contains mistakes and/or took a long time to complete. And, if your goal is automatic performance leading to a reputable product, then process assessment is essential to bring this about.

Given what we have said so far about procedural knowledge, it may appear obvious that procedural knowledge assessment concerns itself with process assessment. While this is true for the most part, there are occasions when the distinction between a process and a product isn't clear. Is a dance,
speech, gymnastics routine, comedy routine, or dramatic performance a process or a product? We will discuss ways to assess products as part of process assessments in the next chapter. In this chapter, however, we'll confine our presentation to ways of assessing processes.

Designing the Assessment Context

As you begin the process of designing an assessment to evaluate your learner's procedural knowledge of a domain-specific basic skill, reflect on what we have said so far about this type of knowledge.

- It is knowledge that shows what a learner can do.
- It should be executed smoothly, efficiently, and automatically.
- It is acquired through stages.
- It is constructed by learners from contexts or situations that are important to them and to you.

Having reflected on the above considerations, follow the steps below to design a process assessment. The applications along the way will help you complete each step.

**Step 1: Clarify the procedure and the reasons for assessing it.**

When you decide to use a procedural assessment, you are making a substantial commitment of your time and energy and that of your learners. You want to have a clear reason for doing so. Ask yourself why it is important for you to assess how someone goes about doing the task instead of just examining and rating the final product? What will you do differently with the results of a procedural assessment in comparison to the results of a product assessment? Some possible reasons include:
• **Individual diagnosis:** you want to observe the strengths and needs of each learner as they carry out the procedure on their way to developing automatic performance;
  
  • **Group needs assessment:** you want to determine the strengths and needs of the entire class and use this information to group learners according to their level of mastery
  
  • **Grading:** you want to base a learner's grade on both product and process outcomes;
  
  • **Certification:** you want to verify that individual learners have mastered certain procedures before moving on to other more complex action sequences;

**EXAMPLES:**

(1) We will conduct an assessment of how the learner uses an electronic calculator in order to diagnose individual strengths and needs.

(2) We will conduct an assessment of the steps involved in long division to allow us to group learners to meet their learning needs.

(3) We will assess how the learner sets up a chemistry experiment in order to include this as part of the chemistry grade.
APPLICATION

Step 1: Clarify the procedure and the reasons for a process assessment.

A. The procedure I will be assessing is:


B. My reason(s) for assessing this procedure are:


STEP 2. Describe the assessment context.

Specify the setting and level of obtrusiveness. Decide whether you will use a natural, structured, or simulated situation for observing the
learner's performance. A natural setting is one in which process assessment occurs without the observer intervening or interrupting the normal flow of events. One way to do this is to embed the assessment in a lesson. The standard of authenticity, which we discussed in Chapter 1, would have the assessments embedded in the lessons that teach the skills. Since practice is a necessary feature of any lesson that is designed to teach a procedure, assessment should occur during practice. For example, the most opportune time to assess the processes involved in long division, oral reading, focusing a microscope, using a balance scale, or booting a floppy disc would be as the learner is practicing these skills.

However, there are occasions when it is not practical or feasible to use natural classroom settings as the assessment context. At these times you may tell the learner that you want to observe him in a particular place and at a certain time to perform a procedure. You ask the learner to do something rather than wait for it to happen.

You do this because the process may be one which is best assessed in a private context that minimizes outside distractions. The learner may be anxious about being assessed when peers are around. Lesson embedded assessment may detract from a teacher's ability to monitor class behavior and create problems for classroom management. This may be especially true with large classes.

In choosing between natural and structured situations the issue of best versus typical performance arises. Do you want to assess what the learner typically does when she reads, types, measures, calculates, plays the trombone, or speaks? Or, do you want to know what she can do when she is consciously trying to do her best? Natural situations tend to elicit the former performance and structured settings the latter.
If the procedure being taught is one that naturally occurs outside a classroom context, like job interviewing, speaking before the school board, ordering food in a restaurant, or planning a route for a two-week backpacking adventure, then you may want to use a simulated situation to observe the procedure which you taught. Again, a desire to make the assessment an authentic one would place a premium on putting the learner in a realistic context.

You will also need to decide if the learners will know that they are being assessed. Obviously, with structured situations that occur outside the natural rhythm of the classroom lesson, learners are aware that they are being assessed. But this is not always the case in natural situations. So you must decide whether or not to tell the students that their oral reading, or calculator skills will be rated.

In making this decision, you should weigh considerations of anxiety and motivation. Telling the student that their performance will be assessed likely enhances motivation and allows you to judge maximal performance. But, this may create anxiety in some learners which limits them from doing their best. In such a situation you may want to assess unobtrusively. As you can see, the issue of anxiety relates directly to reliability and the importance of reducing measurement error or bias.

EXAMPLES:

(1) Our oral reading assessments will be done in a private area of the classroom and the students will know that they are being assessed and that this will be part of their reading grade.

(2) Computer literacy will be assessed during the lesson and the learners will not be aware that we are rating their performance.
(3) We will tell the cosmetology students that their styling skills will be assessed on Friday.

*Specify who will observe the performance.* Typically teachers assess procedures. But the learners can self-assess, peers can rate one another's skill, another teacher or aide can assess, or any combination of the above.

Consider the teacher or other expert observer when:

- The procedure is one that requires the specialized knowledge or skill that only an expert possesses;
- the assessment will be used for high-stakes purposes like a final course grade, or to decide the winner of a competition;

Consider peer or self-ratings when:

- The learners themselves are capable of mastering and applying the scoring criteria;
- time constraints prevent you from observing the procedures and giving the learners timely feedback;
- the learners themselves have nothing to gain from artificially inflating or deflating their performance;

Consider using more than one rater when:

- The assessment will be used for high-stakes purposes and the reliability of the ratings is of critical importance;
- the performance is such that a stronger element of subjective judgment is involved than just a yes/no, did or did-not-do-it decision. We will discuss this point further when we examine how to construct a process assessment instrument.
EXAMPLES:

(1) Learners will rate one another on the quality of their oral class presentations;
(2) The shop teacher will assess the use of proper safety procedures when working with power tools;
(3) With a checklist learners will self-evaluate their use of proper procedures for booting a floppy disc and shutting down the computer system.

**Specify how much evidence is needed.** Here you have several choices:

- Collect one sample of performance on one occasion
- Collect more than one sample of performance on one occasion
- Collect more than one sample of performance on different occasions.

The least time consuming choice is the first one. But remember that learner's have good and bad days just like us. How critical is it that your assessment be a reliable and valid one? Consider:

(1) **The importance of the decision.** If the skill is one that places a premium on safety, or the equipment being used is expensive and easily damaged, then more than one sample of behavior may be needed to be certain that the learner possesses the skill;
(2) **The context.** If the skill is one that is used in a variety of contexts each of which can affect differently the learner's performance, you will have to assess the behavior in these different contexts. Oral presentations before familiar and
unfamiliar peers, familiar and unfamiliar adults, can elicit different types of performance. Assessment in only one of these situations is unlikely to indicate how the learner will perform under different conditions.

(3) **Time constraints.** If your assessment time is limited, you simply may not have time to observe more than one performance.

**EXAMPLES**

(1) Use of safe procedures when working with power tools will be observed three times on three different days.

(2) Proper use of the microscope will be observed for two separate slide set-ups during one class.

**APPLICATION**

Step 2: Describe the assessment context.

A. Setting and level of obtrusiveness:

Natural

________________________________________________________________________

________________________________________________________________________

Structured

________________________________________________________________________

________________________________________________________________________
Step 3. List important behaviors and characteristics of the performance including common errors and mistakes. If you are assessing physical movements (e.g., using a microscope; booting a disc) specify the observable behaviors that must occur to skillfully execute the procedure. Word the statements that designate the behaviors in such a way that they can be answered yes or no; present or absent. Try to make the list of behaviors as inclusive as possible without being too lengthy.

In making the list, don't include behaviors that all students perform without a need for instruction. For example, when setting up a microscope every learner will "hold the slide between their thumb and forefinger." So don't include this behavior on your list. In addition, try to identify behaviors that are clearly inappropriate and for which feedback must be given if they
occur. In making an oral presentation, for example, it is important not to use sarcasm when answering a question.

Some processes like oral performances can't always be defined in terms of clearly observable behaviors that lend themselves to yes/no kinds of judgments. In these cases try as best as possible to identify characteristics which have an observable quality to them. For example, it is important that a student articulate clearly during an oral presentation. But the statement, "articulates clearly," doesn't lend itself to a yes/no decision as readily as a statement like "wipes the slide with lens paper." We will discuss how to score these statements in the next section.

Be as analytical as you can when identifying the attributes of a skilled performance. Resist the temptation to say, "I'll know a good oral presentation when I see one." Identifying the critical attributes of a procedure in observable terms in advance is the single most important thing you can do to insure the quality of your assessment. It will focus your observations on important aspects of the procedure that you might otherwise ignore had you not gone through this step.

EXAMPLES

Using an electronic calculator
Making a presentation using audio-video technology
Subtraction with borrowing
Using a balance scale
Constructing a series circuit with switch, battery, bulb
Speaking before the class
Performing a dance routine
Step 3. List the critical behaviors or attributes of the performance


Step 4. **Design the procedure rating plan.** Choose a scoring system best suited for the purpose of assessment identified in Step 1 and for the process you wish to assess. You have four options: checklists, rating scales, holistic scoring and anecdotal records. Each has certain strengths and limitations and are more or less suitable for different types of purposes and processes.
Checklists. Checklists are suitable for processes that can be described in terms of discrete, observable behaviors as in the examples above dealing with calculator and balance scale use. With a checklist, the observer need only observe the process and mark on an observation form whether the behavior occurred or not. Checklists are efficient. They can be scored while the behavior is taking place and, because they require simple judgments, a relatively large number of attributes can be rated. They are useful for providing diagnostic information to learners. Their principal drawback is that not all procedures lend themselves to such a detailed, observable behavior analysis.

Rating scales. Rating scales are suitable for processes that can't be neatly broken down into behaviors that can be scored as present or absent. The examples given above for making a presentation and giving a speech are best assessed with a rating scale. The most common type of rating scale assigns values to the components of the procedure on a continuum which indicates how often the behavior occurs or its quality. These scales can take a variety of forms:

(1) Uses proper safety procedures in the carpentry shop.
   Never Seldom Occasionally Often Always

(2) Uses appropriate voice volume when addressing the class.
   Never Seldom Occasionally Usually Always

(3) Rate the quality of the students introduction of the presentation.
   Inadequate ______ ______ ______ ______ ______ ______ Superb
(4) Rate the efficiency of the subtraction routine.

Hesitant _______ _______ _______ Smooth

(5) Holds the classes attention while speaking.

Inadequate _______ _______ _______ Adequate

(6) Is sensitive to the needs of the audience.

Unaware of audience _______ _______ _______ Attuned to Audience

(7) Persuasiveness

Lacks enthusiasm _______ _______ _______ Convincing

The job of the rater is to focus on the various aspects of the performance and make a judgment along some qualitative dimension. The judgment is made by simply circling the number or word on the scale which best corresponds to the observer's judgment or placing an "X" on a line between bipolar opposites indicating the proximity of the answer to either pole of the continuum. This judgment often can be made after the procedure or performance has been executed. Rating scales are more time consuming to score than checklists and this limits the number of critical attributes which you can practically assess.

**Holistic scoring.** Holistic scoring is used when the rater observes the procedure, noting the different attributes as defined in Step 3 and their quality, and then assigns some numerical value to the overall quality. Holistic
scoring gives an overall index of performance and is appropriate when the purpose of the assessment that you identified in Step 1, is for grading purposes. It is less time consuming than checklists and rating scales. These latter techniques require many individual judgments while holistic scoring requires only one. However, holistic scoring cannot serve a diagnostic purpose since specific strengths and weaknesses are not noted.

The most common examples of the holistic scoring of performances occur in diving competition, gymnastics, band recitals, and dances. Generally numbers are assigned on a 5 or 10 point scale with decimals being used to indicate finer gradations of performance.

Checklists, rating scales, and holistic scoring are the most efficient assessment techniques for large classes and for procedures that have many attributes. They provide a quick, accurate and easily interpretable means of recording data.

**Anecdotal records.** An anecdotal record is a written account of a procedure that describes what the learner did, and makes evaluative judgments of the performance of the critical attributes. Such a record, as shown in Figure 6.1, does not include numerical values but it does provide a richness of detail and information that is often missing in a checklist or rating scale. While it is more cumbersome than these techniques, it lends itself to assessments of improvement when done on several different occasions. It is time consuming to write an anecdotal record and is best for classrooms with small numbers of learners and for procedures that are relatively simple.

Insert Figure 6.1 about here
APPLICATION

Step 4. Design the procedure rating plan.

A. Describe the checklist

B. Describe the rating scale

C. Describe the holistic scoring procedure
D. Describe the anecdotal record


Step 5. Arrange the behaviors or characteristics in an appropriate format. If you are using a checklist or rating scale, organize the behaviors or characteristics in the sequence that is likely to be followed by the learner and encountered by the observer. The list should be easily readable and contain space where checks can be made, numbers circled, time recorded, frequencies counted, and summate ratings computed if such ratings fit the purpose of the instrument. Once the checklist or rating scale is completed, the instrument should be interpreted based on the purpose(s) specified in Step 1. Examples of a checklist and rating scale are provided in Figures 6.2 and 6.3.

Insert Figures 6.2 & 6.3 about here

APPLICATION

6-28
Step 5. Arrange the behaviors or characteristics in an appropriate format. 
Construct the final form in the box below:

Summary
In this chapter you learned a method for assessing how well learners perform sequences of actions or behaviors across a variety of subject areas. We referred to these procedures as domain-specific basic skills. Your goal in teaching these skills is that learners will be able to perform them skillfully and automatically. The techniques covered in this chapter, if used at regular intervals, will allow you to assess learner development across the cognitive, associative and autonomous stages of basic skill development.

These techniques can be used by themselves to assess the development of basic skill procedural knowledge or in combination with other techniques to assess complex performances that show deep
understanding in a subject area. These performances can involve projects or demonstrations that incorporate processes, written products, or other activities whose purpose is to allow learners the opportunity to demonstrate high levels of thinking and problem solving. We will examine this type of learning and the techniques to assess it in the next chapter.

Activities

1. Using an example of a unit from your teaching field, illustrate how procedural knowledge develops from the cognitive stage, through the association stage, to the autonomous stage.

2. For the example above, describe some domain-specific skills, domain specific strategies, and domain-general strategies that a learner might need in order to achieve the goals of your unit.

3. Describe the assessment context for the procedural knowledge you will teach in the above example with respect to: The number of samples of performance you will observe, the specific behaviors you will observe, and the type of scale you will use for assessing the behavior.

4. Provide a draft of the scale you will use to measure the procedural knowledge in Question 3. Be sure to provide a sufficient number of items to measure the behavior reliably and definitions of the behaviors to be assessed.

Suggested Reading


6-30
A comprehensive treatment of cognitive learning theory and cognitive learning outcomes. It's discussion of procedural knowledge and how it is acquired is especially illuminating.
Chapter 7
Assessing Problem Solving Strategies

In Chapter 1 we defined cognitive strategies as ways of thinking that help a student learn. We pointed out that cognitive strategies go beyond the processes that are naturally required to learn or carry out a task. For example, when writing an essay you naturally spell correctly, form complete sentences, and link sentences and paragraphs with transition words. These are not strategies. A writing strategy would involve consciously cueing yourself to identify a goal for writing, asking yourself whether something that you wrote could be said better, or weighing whether the words you use communicate to a particular audience.

Cognitive learning specialists believe that explicit strategy instruction, along with the teaching of a knowledge base, is essential for helping learners become good thinkers. Their research on the importance of explicit strategy instruction is reflected in almost every educational reform movement of the past two decades. The National Education Goals Panel in their report entitled, "Building a Nation of Thinkers" (National Educational Goals Panel, 1993), urges American schools to give learners skills to learn on their own and to think deeply about challenging subject matter.

Cognitive psychologists and educators use a variety of terms and expressions to describe these cognitive strategies. They use such expressions as "reasoning strategies," "problem-solving methods," "decision-making skills," "higher-order thinking skills," "metacognitive learning capabilities," "domain-specific and domain-general learning strategies," "critical-thinking skills," and "strategic reasoning."

Cognitive strategies are types of procedural or know-how knowledge.
In the previous chapter we discussed procedural knowledge as that type of knowledge which leads to actions or doing things. The procedural knowledge we studied was knowledge that learners have about how to ride a bike, drive a car, focus a microscope, tie their shoes, deliver a speech, or use a balance scale. The procedural knowledge that you will learn about in this chapter is knowledge about how to go about solving problems. We will refer to this type of procedural knowledge as "problem-solving strategies."

Cognitive strategies are used for a variety of accomplishments not just problem solving. They are used to weigh alternative choices and help make decisions, hence the term "decision-making strategies." They are used to critique literature or scientific experiments and are sometimes called, "critical-thinking skills." Some educators make a distinction between decision-making, critical thinking, and problem solving strategies (Crowl, Kaminsky, and Podell, 1997). They believe that problem-solving strategies should be used with narrowly defined problems that have one or two specific solutions, while decision-making and critical-thinking skills should be used when the problem confronting the learner is ambiguous, open-ended, and has multiple solutions.

In this chapter, we consider decision-making and critical thinking to be types of problem solving regardless of the complexity of the problem. Consequently, when we use the term "problem-solving strategies" we encompass thinking and reasoning skills used to solve problems that teachers present to learners in class. Sometimes these problems are well defined as in third or fourth grade mathematics word problems. At other times the problems confronting learners are more unstructured as when 7th graders are asked to work out the details involved in planning a cross country trip, or 10th graders are asked to come up with solutions for coping with global warming.

We focus on problem solving for a number of reasons. Teachers and
educators consider problem solving to be one of the most important goals of instruction. In addition, many recent curriculum reforms recommend that teachers approach classroom learning from a problem-solving perspective. Thus, many language arts teachers teach writing as a form of problem solving, or teach reading comprehension from a problem solving perspective.

In this chapter we will limit our focus to assessing that type of procedural knowledge that makes learners facile problem-solvers. Basically, when challenged by a problem, this knowledge let's learners know how to understand or form a representation of the problem, search the problem space for ways to solve it, gather or collect relevant information, and monitor or evaluate their progress towards finding an answer or solution. The goal of this chapter is to develop your knowledge base about problem-solving strategies and to give you the procedural knowledge about how to assess them.

Who Will Defend The Country: An Illustration

Imagine that you are the president of a small, peace loving, independent nation that sits next to an aggressive country itching to expand its borders. Your country is unprepared for military conflict. So, somehow, you must outwit your belligerent neighbor. You must choose someone as a national security advisor. It just so happens that the reigning world chess champion is a citizen of your nation. Since she is a person of gifted intelligence who possesses superior tactical and strategic problem solving ability, you consider choosing her as your national security advisor. Whom you select has a lot to do with your beliefs about expertise. Do you think that an expert is someone of superior intelligence and problem solving abilities who can transfer these strengths to any domain? If so, you might be inclined to choose the chess champion. On the other hand, you may believe that expertise is
domain-specific and that only someone with extensive training and education in a specific field and years of experience can ever be considered an expert. In this case, you might be highly skeptical of the chess champion's ability to protect your country.

This same debate about general or specific expertise has been going on in American education for a century. At certain times the generalists held sway while at others the specialists were ascendant. And, the debate has important implications for the teaching and assessment of problem-solving strategies. Let's review the evolution of this controversy so as to better appreciate current thinking on the extent to which expertise is a matter of domain-general or domain-specific procedural knowledge.

As we turned to the twentieth century, the view that certain subjects exercised and strengthened the mind like a universal weight machine exercises and strengthens muscles was prominent. In particular, the study of Latin, Greek, Aristotelian logic, and the classics was viewed as essential to develop deep thinkers. Thorndike, an eminent educational psychologist, systematically and convincingly debunked this theory during the 1910's and 1920's. Thorndike was one of the early behaviorists and his work so influenced educators that for decades they abandoned any serious study of the mind and how it works to promote learning.

But, this changed in the early 1970's with the advent of computers and a new generation of educational psychologists interested in studying how the mind works. Their initial work supported the notion that general reasoning and problem solving skills lie at the heart of intelligence and expertise. Had you been familiar with their work, at that time, you might have chosen the chess champion as your national security advisor.

However, their attempts to teach general cognitive strategies to learners
in one context and have them transfer this knowledge to different contexts failed. They discovered that learning strategies to improve your memory for foreign language vocabulary didn't improve your memory for history facts or even English vocabulary. Mastering such problem solving strategies as means-end analysis, working forward, brainstorming, or reasoning by analogy helped learners solve familiar problems but not novel ones.

Thus, by the late 1970's and early 1980's, the generality-specificity pendulum took a pronounced swing to the specificity side. Cognitive psychologists were becoming convinced that expertise was not due to general thinking skills. In addition, they were skeptical of study skills classes in junior or senior high-schools that taught things like note-taking, outlining, how to listen to a lecture, or underlining. Their research suggested that such general learning strategy classes didn't improve listening or reading comprehension or memory beyond simply rereading the material several times over. During that time our chess champion would have been told to stick to what she knows best.

But, in the 1980's cognitive psychologists developed an interesting research methodology called the "expert-novice" study. They would give an expert in a specific area and a novice a problem to solve or task to master and, then, ask them to describe out loud or in writing what they were thinking as they engaged in the activity. Thus, they could compare the thinking of a historian or physicist as they were problem solving with that of an amateur. They also gave experts problems to solve or tasks to master outside their area of expertise and told them to describe their thought processes. Out of this research came several important findings.

It was confirmed that experts employ a specific set of problem solving strategies when working in their areas of expertise. Their ability to flexibly
use these strategies hinged on a well organized and elaborated domain-specific knowledge base. In addition, experts who had excellent recall for facts in their field didn't necessarily remember information outside their field better than novices. Chess champions whose memory for chess positions was remarkable had memories for vocabulary words or strings of numbers that was unremarkable. Experts weren't much better at problem solving outside their field than novices to that field. After reading these findings, one could conclude that our chess champion would be no better at defending the country than any person picked at random off the street.

But, this research also turned up something unexpected. It seems that some novices learn unfamiliar material more quickly than others regardless of the knowledge domain. In other words, whether they are studying unfamiliar material about astronomy, chess, hockey, photovoltaics, or Mesopotamian civilization, some novices acquire information faster than others. Regardless of what they read, these intelligent novices are aware when they don't understand something, know which parts of the text are more difficult for them, are aware of the differences between memorizing something and understanding something and change strategies depending on their goal (Bruer, 1993). In other words, some novices were more aware of themselves as learners and more aware of their thought processes. They were using domain-general or domain independent skills to master new fields.

Cognitive psychologists used the term metacognition to describe these mental competencies. Basically these competencies encompass the skills of knowing when you do and do not understand something, knowing when to use a learning strategy to improve your understanding, being aware of whether you are experiencing more or less difficulty learning as you try out a particular strategy, and being able to compare the effects different strategies
are having on your learning. Cognitive psychologists found out that these
domain-general skills could be taught to school-age children and that they
actually improved their ability to learn in any academic area (Pressley, 1995).

Thus, as we approached the 1990's and continuing through today, a
synthesis has come about in the debate over the significance of domain-
general or domain-specific procedural knowledge. This new perspective
holds that domain-specific procedural knowledge, domain-general knowledge
and metacognition are all important for knowing how to learn in any given
domain. Thus, if you are a history or math teacher, there are three types of
procedural knowledge which you can teach your learners to help them become
better thinkers and problem solvers: teach them (1) problem-solving
strategies that are unique to the field of math or history; (2) general cognitive
strategies to improve their recall of information or comprehension of text; and
(3) metacognitive skills that will remind them to use both general and specific
strategies and monitor their effectiveness as they are learning.

And, so, how do we resolve the problem posed at the start of this
section? Our current knowledge about problem solving leads us to conclude
that the chess champion's detailed knowledge about chess and chess strategies
won't help her in the field of international diplomacy. In other words, if all
she knows about is chess, she won't make a good national security advisor.
However, if she knows how to learn new things, if she has the metacognitive
knowledge to be a rapid learner, if she is aware of how to monitor her
learning, then with a crash course on foreign affairs and some on-the-job
training she could become an effective diplomat.

A Framework For Thinking About Problem Solving

Figure 7.1 depicts the cognitive components of problem solving that
you can directly teach and assess. We have left out of this figure the construct of intelligence. This is not to imply that it is unimportant. However, there is little agreement about what intelligence is and less agreement over what can be done about it. There is far greater consensus surrounding the meaning of the components of problem solving that we have listed and the susceptibility of these components to teaching, learning, and assessment. Let's look closer at some of them.

Insert Figure 7.1 about here

**Knowledge base.** There is almost unanimous agreement among cognitive psychologists that skilled problem-solving in any domain depends on an extensive, elaborated, and well organized knowledge base. Chapter 5 presented ways to assess this.

**Domain-general procedural knowledge.** These include strategies like recognizing when a problem exists, identifying the elements of a problem, searching your memory for what you know, brainstorming, forming images of the problem, etc. Students can learn these, and they can improve their problem solving across many domains, provided they also receive instruction in metacognition.

**Metacognition.** Students must (1) know when to use strategies (whether domain-general or domain-specific), (2) monitor their effectiveness, (3) make judgments about the relative effectiveness of different strategies, and (4) make a commitment to using the more effective strategies in future
problem solving. They must be aware of themselves as problem solvers, planners, decision makers, monitors, and evaluators. Without metacognition, strategy use does not generalize to novel contexts. Most cognitive psychologists treat metacognition as a type of procedural knowledge.

**Domain-specific procedural knowledge.** Each knowledge domain has strategies specific to it that learners must master to become good problem solvers. There are strategies specific to good reading, writing, math problem solving, historical analysis, reasoning about science, and literary criticism.

In the remainder of this chapter you will learn how to assess domain-general and domain-specific procedural knowledge of problem-solving strategies and metacognition. Although we have treated them separately, we will present assessment techniques for domain-general strategies and metacognition together under the heading "general problem-solving strategies." This is because both of these types of procedural knowledge are intended to help students problem solve in any academic discipline.

**ASSESSING DOMAIN-GENERAL PROBLEM SOLVING STRATEGIES**

Some General Principles For Assessing Problem Solving

There are certain general rules which need to guide your development of measurement instruments for assessing problem solving. These apply whether you are assessing domain-general or domain-specific problem-solving skills. We have covered some of them in our discussion of validity and reliability in Chapter 4. To these we add the following:

**Novelty.** The tasks that you give your learner's to solve, the situations that you ask them to react to, should be new and unfamiliar.
Otherwise, problems that appear to demand complex reasoning of learners may only be eliciting rehearsal or recall of a routine learned in class. For example, analysis is a problem solving strategy that learners must be proficient at in order to solve a variety of problems. This strategy typically is taught by explanation, modeling, and practice. But, an assessment that asks them to model an analysis that was practiced in class is only measuring recall.

**Authenticity.** Recall what has been said in previous chapters about the importance of learners constructing their own knowledge bases and memory structures. The same applies to problem-solving strategies. You want to challenge them with problems that grab their attention and make them want to find a solution. In other words you want their problem solving to be goal-directed and not done simply to please the teacher. The best way to ensure that learners construct their own solutions to problems and learn problem solving strategies is to make the problem-solving tasks ones that real writers, geographers, historians, mathematicians, literary critics, or scientists face.

**Sustained, purposeful problem-solving.** The assessment task should require prolonged effort and thought. A problem that can be solved in a few minutes is probably not demanding enough of the problem solving skills of your learners. A math problem should require several problem solving strategies rather than a listing of a series of required computations. A problem built around an historical event should ask learners to analyze and interpret, and not just identify some key point.

**Problems with multiple solutions.** In the real world, we rarely encounter problems that have one solution. When writing something for a
newspaper, there are a variety of ways to approach the task. The good writer makes intelligent choices about the best way to reach his or her audience. Likewise there are a variety of locations to build a city, or a railroad, locate a solid-waste disposal plant, plan an exploration, increase voter registration, etc. Thus, the problems you ask learners to solve should have multiple-solutions and require the learner to identify multiple alternatives and their relative merits.

**Evidence of the problem-solving process.** The format that you design should require not just an answer but an explanation of it. Learners must be required to explain or show how they arrived at the answer. This is the only way that you can judge if they have learned the problem-solving strategy. Open-ended question formats rather than objective-type formats are best for eliciting this type of response.

**Ratings.** In the last chapter we introduced the idea of ratings to judge the quality of procedures. Likewise assessing problem-solving strategies will require skillfully designed rating procedures. Such assessment by its very nature requires judgment. No true-false, completion, or multiple-choice item will elicit real problem-solving from your learners. You will have to make a judgment about how well they reason. This judgment has a strong subjective element. It can become more objective through the use of rating procedures.

A Taxonomy Of General Problem-Solving Strategies

Since the late 1960's when many in the field of educational psychology turned away from behaviorism and renewed their interest in how the mind works, psychologists have developed systems or taxonomies for describing
reasoning and problem-solving. Bloom's et al.'s *Taxonomy of Educational Objectives*, reviewed in Chapter 3, was one of the first and most influential examples of these classification systems. As you will recall, they classified the cognitive domain into six categories of learning: Knowledge and Understanding (which represent lower level reasoning), Application, Analysis, Synthesis, and Evaluation (which represent increasingly higher levels of mental effort).

Ellen Gagne (Gagne, Yekovich & Yekovich, 1993) defines problem solving as a situation where one has a goal and hasn't yet figured out a way to reach it. The goal may be needing to answer a question at the end of a history chapter, solving a theorem in geometry, or writing a letter to the editor of the local newspaper. In each situation there is a goal that has not been reached. Problem solving for Gagne involves the following cognitive processes regardless of the specific nature of the problem:

1. **Forming a representation of the problem.**
   The learner must think about the information given and what she is being asked to do.

2. **Searching the problem space.**
   The learner searches her memory for what she knows about the problem and different ways to solve it.

3. **Choosing a solution.**
   The learner applies her knowledge to solve the problem.

4. **Evaluation.**
   The learner judges whether her application worked.

Ennis (Ennis, 1987) classifies problem solving into skill clusters that
involve:

(1) Clarifying issues and terms,
(2) analyzing and identifying the components of an argument,
(3) judging the credibility of evidence,
(4) using inductive and deductive reasoning,
(5) handling fallacies, and
(6) making value judgments.

For Ennis, problem solving involves reflective thinking that is focused on deciding what to do about a goal, what to believe, and what to do.

Quellmalz has developed the most thoroughly researched general problem solving taxonomy and has employed it over the past decade in numerous curriculum and assessment programs (Quellmalz, 1987; 1991; Quellmalz and Hoskyn, 1997). Quellmalz classifies problem solving into four categories of reasoning strategies: analysis, comparison, inference and interpretation, and evaluation. She states that problem solving involves the deliberate use of these fundamental cognitive strategies. These strategies are non-hierarchical, unlike those proposed by Bloom. He has also identified four categories of metacognitive strategies that learners apply when they are problem solving: planning, draft and tryout, monitor and revise, evaluate and reflect.

Quellmalz stipulates that these cognitive and metacognitive operations are carried out in all problem solving regardless of the content area. However, he also underscores that learners apply these processes in specific contexts like biology, history, or writing. Consequently, the best way to teach general problem solving is by presenting significant, authentic, and meaningful
problems to learners in specific academic disciplines. In the remainder of this section we will demonstrate how to assess general problem solving strategies using Quellmalz's taxonomy. But, before doing so, we will briefly address how to teach them.

Teaching Problem Solving Strategies

As we have stressed in this chapter and previous ones, it is important that learners construct problem solving strategies for themselves; they should not be spoon-fed in some rote learning fashion. Nevertheless, research on the teaching of problem solving strategies underscores that the teacher plays a direct role in helping learners construct them. He/she does this by giving direct explanations of the strategies and allowing learners to actively learn from them. Pressley refers to this as "Direct Explanation Teaching" (Pressley, 1995).

The basic features of direct explanation teaching of general (or domain-specific) problem solving strategies are the following (Borich & Tombari, 1997):

**Up-front demonstration and explanation.** The teacher tells her learners the goal of the problem solving lesson, what is expected of the learners, presents a significant task that requires the use of a problem-solving strategy, and informs the learners that they will need to learn the strategy to solve the problem.

**Mental modeling.** The teacher thinks the strategy aloud and, as they watch and listen, shows the learners how to use the strategy.
**Guided practice** The teacher provides an authentic task and uses hints, prompts, and questions to get learners to use the modeled strategy. Feedback, praise, and encouragement accompany the learners' efforts.

**Metacognitive information.** The teacher points out to her learners when and where to use the problem solving strategy and how to notice or monitor whether it is helping them.

Once you have made a commitment to incorporate general reasoning strategies into your curriculum, then the following steps should be followed to assess them:
Step 1. Identify and Define the General Problem Solving Strategy(ies)

Before you can teach problem solving and assess how well your students have learned the strategies, you first need to be clear about what the strategies are. We will use Quellmalz's taxonomy as a model for the assessment of general problem solving strategies. Quellmalz advocates teaching all the components of problem solving and metacognition. He also stresses that one component is not a prerequisite for learning another. You can teach these components in any particular order and you can select some and not others. Quellmalz's work in public schools has involved teaching learners all the component strategies that are involved in problem solving. They are identified and defined as follows:

**Cognitive strategies.**

*Analysis:* analysis is a strategy that involves perceiving the whole task or problem; identifying the distinctive elements, aspects, or terms associated with the problem; understanding the relationship of each distinctive element to the whole task.

*Comparison:* comparison is a strategy that involves identifying the similarities and differences among the various elements or aspects of some task, choosing the dimensions along which to make comparisons or contrasts, and understanding the significance of the comparison for solving the problem presented.

*Inference and interpretation.* Inference and interpretation involve gathering information or evidence to solve a problem, using inductive or deductive reasoning and applying certain rules or heuristics to reach a conclusion.

*Evaluation.* Evaluation requires learners to take a position or arrive at
a conclusion, identify certain criteria for judging the adequacy of the
government or position, presenting evidence of how well the solution or
position meets or does not need the various criteria, and coming to
some judgment about the adequacy of the conclusion or solution.

**Metacognitive Strategies**

*Planning.* Planning occurs when learners analyze the problem,
compare the problem to problems they encountered previously, and
identify strategies to solve the problem.

*Draft and Tryout.* Involves several series of attempts to solve the
problem.

*Monitor and Revise.* The learner makes preliminary checks on whether
goals are achieved and strategies are working.

*Evaluate and Reflect.* The learner examines the adequacy of the
problem solution and the effectiveness of strategies used to reach a
solution.

The definitions given above are suggestive but by no means inclusive of
all that's involved in analysis, evaluation, planning, etc. You may want to
expand each definition to fit better with how you teach analysis or evaluation
or planning in a particular content area. Or, you may be using a curriculum
that defines these problem solving strategies somewhat differently. The point
is that you must be clear about what each strategy entails before teaching and
assessing them. This step of the assessment process is completed by
identifying what your learners must do to demonstrate the skill you wish to
teach. The *Higher Order Thinking and Problem Solving Checklist* at the end
of this chapter can help you select and prioritize some of the strategies you
may want to teach.

APPLICATION

Identify a particular cognitive or metacognitive problem solving strategy that you wish to teach and describe its dimensions:

Learners demonstrate the skill of ___________ when they do the following things:

1. __________________________________________
2. __________________________________________
3. __________________________________________
4. __________________________________________

We will return to this task in Step 4 when we explain how to rate the quality of the learners' problem solving strategies.

Step 2: Select the Context for Demonstrating One or More of the Above Strategies

Once you have decided which of the above strategies you wish to teach and assess, you must select a context that motivates the student to want to demonstrate knowledge of the strategy. An authentic context is preferable, but a novel context is mandatory. The cognitive strategies of analysis, comparison, inference, and evaluation are challenging to learners when they are asked to use them to solve newly encountered problems. Here are
suggestions for contexts that many teachers use who teach and assess problem solving:

**Cognitive strategies: literature.** Give your learners an unfamiliar narrative story and have them identify the key story elements, settings, plot events, and character traits (analysis). Or, give them a newspaper editorial and have them analyze the content for positions, reasons, evidence and conclusions. Assess comparison by giving several editorials and asking them to compare and contrast points of view, evidence, accuracy, and organization.

**Cognitive strategies: history.** Have your learners read an important historical document such as Lincoln's Gettysburg address, and have them deduce causes and events, predict future effects, and infer consequences. Or, have them read about a particular policy such as the President's tax policy and evaluate its significance, practicality, impact, and lasting effects.

**Metacognitive strategies: science.** You can assess how thoroughly learners plan an experiment to test the effects of climatic change on mountain ecosystems, or how adequately they plan a project to assess the effect of gravity on the speed of a flying object. The problem solving strategies of drafting and trying out can be assessed in an electrical conductance experiment where learners have to determine which of a variety of substances conduct electricity the best.

**Metacognitive Strategies: social science.** You can assess the problem solving strategies of monitoring and revising in a psychology project designed to teach an animal to learn a particular behavior. Your learners can monitor
improvement in learning and make adjustments if their goals are not being met. In the same context, they can evaluate the final outcome of the study and reflect on what they could have done differently.

APPLICATION

In the space below, specify the particular context that you will use to teach and assess the problem solving strategy identified in Step 1.

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

Step 3: Design the Assessment Task

In designing the assessment task you will need to answer two questions: (1) What type of outcome or performance will you observe to assess the learner's problem solving strategies? and (2) What type of instructions will you give the learner for carrying out the task? Let's look at each question.

Process vs product. You can assess problem solving by observing
learners as they are engaged in the activity and rating the extent to which you see evidence of analysis, evaluation, planning, revision, etc. An obvious example would be observing learners as they carry out a science experiment or solve a math problem. Or, you could observe learners as they engage in a mock classroom political debate and assess such reasoning strategies as analysis, evaluation, inference, planning, or reflection. In addition to observing and rating, you could ask questions of your learners that elicit their thinking and reasoning.

But, as we said in the previous chapter, processes are difficult to evaluate. They require on-the-spot judgments and are best suited for performances that can be rated as present/absent, yes/no. Problem-solving strategies can rarely be rated in this manner. They exist more in matters of degree than in absolute dichotomies.

So, the better choice is to design an assessment task that results in some product. The product may be a written essay which you will read and assess for evidence of analysis, comparison, inference, planning, etc. While this type of product often works well for junior and senior high school learners, it doesn't work as well for younger learners. For them, products that combine narrative with more visual or pictorial detail may be more conducive for the assessment of general reasoning strategies. Here are some examples of this type of product.

**Visual maps.** Assess how well learners can analyze the elements of a story by having them draw characters, settings, and events and use arrows to depict sequences. Captions can be included that identify plot elements, episodes, goals, or outcomes. Figure 7.2 is an example of a visual story map for "The Three Little Pigs." Such visual techniques can be used to show

7-21
evidence of the metacognitive skills of planning and revision. Mapping formats, as we saw in Chapter 5, can also be used to assess comparison as shown in Figure 7.3.

Insert Figures 7.2
(Art: Revise to fit story of three pigs)

&

Insert Figure 7.3 about here

-Outlines. Analysis, inference, planning, draft and tryout can be assessed using outline formats.

- Timelines, flowcharts, drawings, graphs, pie charts, genealogical charts. Learners can demonstrate planning by constructing a timeline that depicts important milestones in the process of completing a science project. A genealogical chart can show skill in analyzing a story dealing with generations of family members. Flowcharts can be used to assess analysis, planning, and inference. Students can be asked to construct graphs or pie charts as part of a math problem solving activity. Such graphs or pie charts can be examined for evidence of inference and interpretive problem solving strategies. Following a science unit on bones and joints, learners can be requested to draw a picture of what a particular animal or person would look like who could perform certain kinds of actions like use tools, run exceptionally fast, or climb trees.

- Task instructions. Since you want the product to show evidence of particular reasoning strategies, be sure to include explicit cues to learners to
show evidence of the strategy they are using or the thinking and reasoning they went through on the way to solving the problem. Reminders like, "Show all work," "Be sure to list the steps involved," etc. will allow you better to assess both cognitive and metacognitive strategies.

Quellmalz strongly recommends that, particularly with tasks that assess analysis and comparison, you include the question, "So what?" Rather than simply list elements in an analysis or develop a laundry list of similarities and differences, learners should be asked to explain the significance of the analysis or points of comparison. Thus, task directions should explicitly cue learners to explain why the various aspects of the analysis or comparison are important for the story as a whole (e.g." Why should the reader care about the things you analyzed?" or "Why are the similarities and differences that you discussed important for an understanding of this historical event.")

APPLICATION

In the space below describe the specific task that you will give your learners that will allow them to demonstrate the problem solving strategy that you specified in Step 1. Be sure to include explicit directions for what you want them to do:


7-23
Step 4: Develop the Scoring Procedure

Open-ended or extended essay answers, story maps, diagrams, etc., must be scored according to explicit criteria. These criteria must define the important features of analysis, comparison, planning, etc. that the learner is expected to demonstrate and the range of quality presented in the product. If you took the time necessary to complete Step 1, then you have completed the lion's share of the work involved in this part of the assessment process. Your scoring guide simply lists the elements of your definition of analysis, comparison, etc. and includes a rating scale for each element.

For example, let's say that in Step 1 you listed the following as elements or features of narrative story analysis: Learners use the cognitive strategy of story analysis when they:

1. Name or identify the important elements of a story (plot, character, etc.)
2. Describe, explain, or give detail about each element
3. Explain how each element is important for the story as a whole
4. Organize the analysis in some logical fashion

Then your scoring guide for assessing analysis might look something like this:
Assessing Analysis

1. Names/defines the important elements of the story

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lists some elements</td>
<td>Names all elements</td>
<td>and defines some</td>
<td>Names/defines all</td>
<td>elements</td>
</tr>
</tbody>
</table>

2. Gives details about each element.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lacks detail</td>
<td>Adequate detail</td>
<td></td>
<td>Abundant detail</td>
<td></td>
</tr>
</tbody>
</table>

3. Explains how each element is important for the story.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>Some explanation</td>
<td></td>
<td>Full explanations</td>
<td></td>
</tr>
</tbody>
</table>

4. Organization

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disorganized</td>
<td>Some organization</td>
<td></td>
<td>Logical presentation</td>
<td></td>
</tr>
</tbody>
</table>

The points on the rating scale and the terms associated with them define the range of quality of each feature or aspect of the cognitive strategy that you wish to assess. Rating scales are typically constructed with 3, 5, or 7 point
Similarly, your guide for assessing analysis might take the following form:

1. Identifies important aspects of the product to be evaluated
2. Identifies and explains criteria for evaluation
3. Presents evidence on how well each aspect meets the criteria
4. Draws a conclusion about the adequacy of the product

You would then construct a 3, 5, or 7 point rating scale for each part of the strategy and define the end and middle points of the scale as was done in the example above.

Thus, for each cognitive and metacognitive problem solving strategy that you wish to assess, you could construct similar rating scales. Your scoring task involves examining the product thoroughly and then circling the number on the rating scale which best reflects your judgment of how well it meets the criteria. You then add up the total points earned by the student and associate this total with a grade. We will discuss grading practices in Chapter 11.

APPLICATION

In the space provided, list the features of the problem solving strategy that you identified in Step 1 and construct a rating scale for each feature:

1. ________________________________
2. ________________________________
ASSESSING DOMAIN SPECIFIC PROBLEM SOLVING STRATEGIES

General problem solving strategies like analysis or planning can be applied to solving problems in any domain or discipline. We pointed out that recent research shows that many learners are intelligent novices, that is, they have learned these general ways to think, reason, and problem solve that help them learn new things quickly.

But, we also stressed that each knowledge domain—for example, history, physics, reading comprehension, writing, literature, geography, psychology—has specific problem solving strategies associated with it that experts in these domains know and use. In other words, strategies used for solving problems in biology are different from those used to solve problems in political science. There are math problem solving strategies that are unique to particular areas of math like geometry or algebra. These strategies can also be taught and assessed. Let's see how to do this.

Identifying Domain-Specific Problem Solving Strategies

As we pointed out, educational psychologists starting from the time of Bloom, have been developing systems or taxonomies of general problem-solving strategies. We know about specific problem-solving strategies from studying how experts in domains like physics, mathematics, writing, history, or horse race handicapping solve problems. We discussed some of these expert studies in the first section of this chapter.

Thus, there are two basic ways to find out about domain-specific
problem solving strategies: if you are an expert in a particular area, write down exactly how you go about solving problems; or consult textbooks that specifically deal with strategies in the subject areas (see, for example, McGilly, 1993). Here are some examples of domain specific problem-solving strategies:

**Physics.** Experts in solving physics problems dealing with motion and forces do the following:

1. Draw a diagram that depicts all the forces given in the problem,
2. indicate how the forces interact,
3. apply mathematical operations to information given in the problem to generate new knowledge, and
4. review problem question and solve it.

**Reading comprehension.** Expert readers do the following as they are reading texts:

1. Summarize what they have read,
2. ask questions to themselves about ideas they are unsure of,
3. clarify uncertainties, and
4. predict what occurs next.

**Writing.** Expert writers do the following:

1. Plan (search long-term memory for what they know; organize information; set goals),
2. translate (generate text), and
3. review (evaluate and revise their writing).
Historical analysis. Expert historians do the following as they study about past events:

(1) Corroborate evidence from different sources,
(2) consider the credibility of the source, and
(3) contextualize an interpretation.

Solving simple math problems: estimation. Expert mathematicians do the following to solve estimation problems:

(1) Reformulate (round off numbers to make them more manageable),
(2) translate (change the structure of the problem from the one originally given), and
(3) compensate (compensate for one adjustment by making an equal and opposite adjustment).

These are but a few examples of the strategies that domain experts use to solve problems in their areas. As you have realized by now there are as many domain-specific problem solving strategies as there are domains and sub domains of knowledge. The procedures for assessing these strategies are nearly identical to those for assessing general reasoning strategies. Below is an example of how this is done.

Step 1: Identify the Strategy

The strategy to be assessed is the learner's skill in designing a scientific experiment. The steps involved are:

(1) Generates a question
(2) Generates an hypothesis
(3) Makes a prediction
(4) Conducts a controlled experiment to test the prediction
(5) Collects, organizes, and interprets relevant data
(6) Draws reasonable conclusions from the data.

Step 2: Select the Context for Demonstrating the Strategy

The learner will demonstrate the strategy in a physics lab experiment. The problem will be the following: Under normal atmospheric conditions, an object is placed on a scale and the scale reads 10 pounds. If there were no gravity, what would the scale read?

Step 3: Design the Assessment Context

The learner will carry out the experiment keeping a detailed log of all the steps involved in the experiment and the reasoning behind each step. She will show all work and calculations. Drawings will be made of all aspects of the experimental procedure. While the learner is conducting the experiment, the teacher will question her about her reasoning. Both process and product assessment procedures will be used.

Step 4: Design the Scoring Procedure

The following will be assessed based on oral questions and responses:

(1) Explains the reasoning behind hypotheses.
(2) Makes reasonable predictions.
(3) Explains the basis of the prediction.

The following will be assessed based on the written report:

(1) States a question
(2) Conducts a controlled experiment
(3) Collects, organizes, and interprets relevant data.

(4) Draws reasonable conclusions.

Summary

In this chapter you have learned how to identify and assess domain-general and domain-specific problem solving strategies. We want to emphasize two points. First, don't underestimate the amount of time it takes to teach these strategies. Too often we have seen teachers explain and model a problem-solving strategy on one occasion and expect learners to use it from then on. You must not only model it, but also must provide relevant practice and give learners metacognitive knowledge about when to use it and how to monitor its effectiveness. Second, use authentic contexts to teach and assess strategy learning. Learners will only master these strategies and want to show off what they have learned if the contexts for learning and assessment are important ones. The next chapter will focus on developing performance assessments that use real contexts to assess learner acquisition of knowledge, basic procedural skills, and problem-solving strategies.

Activities

1. You have been chosen by your superintendent to interview four candidates for the principal of your school. Compose a list of domain-specific questions and domain-general questions that you will ask each candidate, the answers to which, in your opinion, would tell you the best person for the job. Indicate whether you will place more emphasis on domain-specific (e.g., knowledge of the school and community) or domain-general (e.g., decision-making and problem solving strategies) questions--or whether you will consider them both equally, and why?
2. Assume that after the first round of interviews two of the candidates above are tied for the position of principal. You have been asked to interview both candidates, this time with instructions from the superintendent to determine which candidate is more aware of themselves as a problem-solver and of their own decision-making processes. In other words, the superintendent wants to know which candidate is likely to be more in-tune with the management techniques and strategies they are using and their desired effects and, if need be, flexible enough to change. What questions would you now ask the candidate to provide the information requested?

Suggested Reading

Higher Order Thinking and Problem Solving Checklist

Check each column below indicating (a) the extent to which your curriculum requires students to achieve the following outcomes and (b) the extent to which you are teaching your students to achieve these outcomes.

Assign the number 5 to each checkmark under "Great Extent," a 4 to "Fair Extent," a 3 to "Some Extent," a 2 to "A Little," and a 1 to "Not at All." Subtract your assigned values for the Degree of Implementation column from the Degree of Importance column for each behavior to arrive at your highest priorities.

<table>
<thead>
<tr>
<th>Some Higher Order Thinking and Problem Solving Behaviors</th>
<th>Degree of Importance</th>
<th>Degree of Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Does your curriculum require students to achieve the following:</td>
<td>Are you teaching your students to achieve the following:</td>
</tr>
<tr>
<td></td>
<td>(Check one)</td>
<td>(Check one)</td>
</tr>
<tr>
<td></td>
<td>To a Great Extent</td>
<td>To a Great Extent</td>
</tr>
<tr>
<td></td>
<td>To a Fair Extent</td>
<td>To a Fair Extent</td>
</tr>
<tr>
<td></td>
<td>To Some Extent</td>
<td>To Some Extent</td>
</tr>
<tr>
<td></td>
<td>A Little</td>
<td>A Little</td>
</tr>
<tr>
<td></td>
<td>Not at All</td>
<td>Not at All</td>
</tr>
<tr>
<td>1. Form a mental or symbolic representation of a given problem.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Search his/her long term memory for what is already known about a problem and different ways to solve it.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Analyze and identify the components of an argument.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Judge the credibility of evidence.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Catch fallacies and contradictions.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>To a Great Extent</td>
<td>To a Fair Extent</td>
</tr>
<tr>
<td>---</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>6.</td>
<td>Make value judgments.</td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Identify the similarities and differences among various elements of a problem.</td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>Understand the relationship of each element of a problem to the whole problem.</td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>Gather information or evidence to solve a problem using inductive or deductive reasoning</td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>Apply given rules or heuristics to reach a conclusion.</td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>Identify criteria for judging the adequacy of a position or conclusion.</td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td>Present evidence of how well a solution or position meets various criteria.</td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>Judge the adequacy of a conclusion or solution.</td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>Compare a problem to problems encountered previously</td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>Compose drafts and tryouts in attempts to solve a problem.</td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>Examine the adequacy of strategies used to reach a solution.</td>
<td></td>
</tr>
<tr>
<td>17.</td>
<td>Distinguish the most important elements of a problem.</td>
<td></td>
</tr>
<tr>
<td>18.</td>
<td>Organize a conclusion about a problem in a logical fashion.</td>
<td></td>
</tr>
<tr>
<td>19.</td>
<td>Identify criteria for evaluating a problem solution.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>To a Great Extent</td>
<td>To a Fair Extent</td>
</tr>
<tr>
<td>---</td>
<td>------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>20</td>
<td>Draw a diagram that depicts all the forces impacting a problem.</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Apply given logical or mathematical operations to a problem to generate new knowledge.</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Summarize what is read, orally and written.</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Asks questions to oneself about ideas he/she is unsure of.</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Predict what will occur next in a sequence of events.</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Find corroborating evidence from among different data sources.</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Place an interpretation of a problem in the context of prevailing circumstances.</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Evaluate and revise what is written.</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Reformulate to make a problem more manageable.</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Change the structure of a problem from the one originally given.</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Compensate for an adjustment by making an equal and opposite adjustment.</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Plan/conduct a controlled experiment to test one's prediction.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>32. Explain the reasons or theory behind an hypothesis or prediction.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33. Clearly communicate the results of what was observed in written and oral format.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>To a Great Extent</td>
<td>To a Fair Extent</td>
</tr>
<tr>
<td>---</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>34.</td>
<td>Draw a conceptual map or picture that shows what was learned or observed.</td>
<td></td>
</tr>
<tr>
<td>35.</td>
<td>Consult experts, magazines, encyclopedias, newspapers, scientific journals, etc., for new information.</td>
<td></td>
</tr>
<tr>
<td>36.</td>
<td>Demonstrate independence and autonomy in completing a project or demonstration.</td>
<td></td>
</tr>
<tr>
<td>37.</td>
<td>Construct and interpret graphs charts, and tables.</td>
<td></td>
</tr>
<tr>
<td>38.</td>
<td>Make reasonable conjectures based on analysis of data.</td>
<td></td>
</tr>
<tr>
<td>40.</td>
<td>Accurately summarize what others have said.</td>
<td></td>
</tr>
<tr>
<td>41.</td>
<td>Provide assistance to others when asked.</td>
<td></td>
</tr>
<tr>
<td>42.</td>
<td>Ask for feedback when needed.</td>
<td></td>
</tr>
<tr>
<td>43.</td>
<td>Change viewpoint to match the facts.</td>
<td></td>
</tr>
<tr>
<td>44.</td>
<td>Listen attentively to others.</td>
<td></td>
</tr>
<tr>
<td>45.</td>
<td>Use a constructive tone when responding to others.</td>
<td></td>
</tr>
<tr>
<td>46.</td>
<td>Share and take turns.</td>
<td></td>
</tr>
<tr>
<td>47.</td>
<td>Set goals that are achievable within a specific span of time.</td>
<td></td>
</tr>
<tr>
<td>48.</td>
<td>Consult a variety of knowledge sources to achieve deep understanding.</td>
<td></td>
</tr>
</tbody>
</table>
49. Realistically evaluate own performance.
<table>
<thead>
<tr>
<th></th>
<th>To a Great Extent</th>
<th>To a Fair Extent</th>
<th>To Some Extent</th>
<th>A Little</th>
<th>Not at All</th>
<th>To a Great Extent</th>
<th>To a Fair Extent</th>
<th>To Some Extent</th>
<th>A Little</th>
<th>Not at All</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>50. Keep a record of one's own progress toward important goals.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>51. Ignore distractions that interfere with goal attainment.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>52. Help keep others on task.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

7-39
Figure 7.1. Cognitive Components of Problem-Solving
Higher Order Thinking and Problem Solving Checklist

Check each column below indicating (a) the extent to which your curriculum requires students to achieve the following outcomes and (b) the extent to which you are teaching your students to achieve these outcomes.

Assign the number 5 to each checkmark under "Great Extent," a 4 to "Fair Extent," a 3 to "Some Extent," a 2 to "A Little," and a 1 to "Not at All." Subtract your assigned values for the Degree of Implementation column from the Degree of Importance column for each behavior to arrive at your highest priorities.

<table>
<thead>
<tr>
<th>Some Higher Order Thinking and Problem Solving Behaviors</th>
<th>Degree of Importance</th>
<th>Degree of Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Does your curriculum require students to achieve the following:</td>
<td>Are you teaching your students to achieve the following:</td>
</tr>
<tr>
<td></td>
<td>(Check one)</td>
<td>(Check one)</td>
</tr>
<tr>
<td></td>
<td>To a Great Extent</td>
<td>To a Great Extent</td>
</tr>
<tr>
<td></td>
<td>To a Fair Extent</td>
<td>To a Fair Extent</td>
</tr>
<tr>
<td></td>
<td>To Some Extent</td>
<td>To Some Extent</td>
</tr>
<tr>
<td></td>
<td>A Little</td>
<td>A Little</td>
</tr>
<tr>
<td></td>
<td>Not at All</td>
<td>Not at All</td>
</tr>
</tbody>
</table>

1. Form a mental or symbolic representation of a given problem.

2. Search his/her long term memory for what is already known about a problem and different ways to solve it.

3. Analyze and identify the components of an argument.

4. Judge the credibility of evidence.

5. Catch fallacies and contradictions.
<table>
<thead>
<tr>
<th></th>
<th>To a Great Extent</th>
<th>To a Fair Extent</th>
<th>To Some Extent</th>
<th>A Little</th>
<th>Not at All</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.</td>
<td>Make value judgments.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Identify the similarities and differences among various elements of a problem.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>Understand the relationship of each element of a problem to the whole problem.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>Gather information or evidence to solve a problem using inductive or deductive reasoning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>Apply given rules or heuristics to reach a conclusion.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>Identify criteria for judging the adequacy of a position or conclusion.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td>Present evidence of how well a solution or position meets various criteria.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>Judge the adequacy of a conclusion or solution.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>Compare a problem to problems encountered previously</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>Compose drafts and tryouts in attempts to solve a problem.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>Examine the adequacy of strategies used to reach a solution.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.</td>
<td>Distinguish the most important elements of a problem.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.</td>
<td>Organize a conclusion about a problem in a logical fashion.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19.</td>
<td>Identify criteria for evaluating a problem solution.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>20</td>
<td>Draw a diagram that depicts all the forces impacting a problem.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Apply given logical or mathematical operations to a problem to generate new knowledge.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Summarize what is read, orally and written.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Asks questions to oneself about ideas he/she is unsure of.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Predict what will occur next in a sequence of events.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Find corroborating evidence from among different data sources.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Place an interpretation of a problem in the context of prevailing circumstances.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Evaluate and revise what is written.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Reformulate to make a problem more manageable.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Change the structure of a problem from the one originally given.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Compensate for an adjustment by making an equal and opposite adjustment.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Plan/conduct a controlled experiment to test one's prediction.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Explain the reasons or theory behind an hypothesis or prediction.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
33. Clearly communicate the results of what was observed in written and oral format.
<table>
<thead>
<tr>
<th></th>
<th>To a Great Extent</th>
<th>To a Fair Extent</th>
<th>To Some Extent</th>
<th>A Little</th>
<th>Not at All</th>
<th>To a Great Extent</th>
<th>To a Fair Extent</th>
<th>To Some Extent</th>
<th>A Little</th>
<th>Not at All</th>
</tr>
</thead>
<tbody>
<tr>
<td>34.</td>
<td>Draw a conceptual map or picture that shows what was learned or observed.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35.</td>
<td>Consult experts, magazines, encyclopedias, newspapers, scientific journals, etc., for new information.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36.</td>
<td>Demonstrate independence and autonomy in completing a project or demonstration.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37.</td>
<td>Construct and interpret graphs, charts, and tables.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38.</td>
<td>Make reasonable conjectures based on analysis of data.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40.</td>
<td>Accurately summarize what others have said.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41.</td>
<td>Provide assistance to others when asked.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42.</td>
<td>Ask for feedback when needed.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43.</td>
<td>Change viewpoint to match the facts.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44.</td>
<td>Listen attentively to others.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45.</td>
<td>Use a constructive tone when responding to others.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46.</td>
<td>Share and take turns.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47.</td>
<td>Set goals that are achievable within a specific span of time.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48.</td>
<td>Consult a variety of knowledge sources to achieve deep understanding.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49.</td>
<td>Realistically evaluate own performance.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>To a Great Extent</td>
<td>To a Fair Extent</td>
<td>To Some Extent</td>
<td>A Little</td>
<td>Not at All</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>------------------</td>
<td>------------------</td>
<td>---------------</td>
<td>---------</td>
<td>------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50. Keep a record of one's own progress toward important goals.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51. Ignore distractions that interfere with goal attainment.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52. Help keep others on task.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 8
Assessing Deep Understanding

In Chapter 5 you learned how to construct objective-type questions and simple essays to measure a learner's knowledge. That chapter also gave you information about how to assess simple understanding which we defined as an organized knowledge base. Here are some activities that get at a different type of learning:

The Courier is interested in its readers' opinion of a new city tax that will help offset the cost of a new football stadium for the city's NFL team. The story shown below is based on a poll taken by the Courier of a random sample of 200 readers surveyed on three separate occasions. (The story is about a decline in support of the tax. The decline became even more pronounced when its NFL team failed to make the playoffs). Write a letter to the editor about this story that reflects what you have learned this past month about survey research, statistics and sampling.

Choose a neighborhood from the city map on your table and identify its important features. From the information available to you identify any significant problems like poor housing, crime, traffic congestion, etc. Consider various plans for improving the neighborhood and decide how to implement them. Select one plan and prepare a written and oral demonstration to the city council. Be sure to include any alternate plans and why you rejected them. Your plan must be ready a week from today.
The desks in front of you contains batteries, wires, alligator clips, light bulbs, paper clips, plastic and wooden spoons. Find out which substances make the bulb light and then develop a presentation that explains what a conductor is and why it conducts.

Learners who address these problems correctly have accomplished some significant learning feats. They have acquired a knowledge base in math, geography, or science. This knowledge base undoubtedly has some organization or structure to it that allows learners to quickly recall important concepts, rules, and generalizations. They have learned procedures for calculating math problems, drawing diagrams that represent forces, or constructing maps. In addition, these students have learned certain problem solving strategies like analysis, inference, evaluation, planning, tryout and revision. In other words, these learners draw upon their declarative, procedural, and metacognitive knowledge to demonstrate deep understanding of an academic discipline.

What exactly do we mean when we use the term "deep understanding?" In everyday language, deep understanding means that you "really know" something. And when you "really know" something like physical forces and motion, African history, Spanish literature, or learning theory you have the initiative and commitment to learn new things in these areas and construct new knowledge for yourself. You can organize this new knowledge into concepts and generalizations and see relationships and patterns. Moreover, you can apply this knowledge to solve novel problems.

This chapter is about assessing deep understanding. In order to do this you will use much of what you have learned about assessing declarative and
procedural knowledge, problem solving strategies and metacognition. The term applied to techniques for assessing deep understanding involving the observation and measurement of a student process or product is "performance assessment." In this chapter we hope to clarify exactly what performance assessment is. But before doing so, let's take a critical look at how deep understanding has typically been assessed in schools.

The Essay Exam

You have undoubtedly taken many essay exams by now. They have asked you to do things like "compare and contrast," "discuss thoroughly," "take a position and defend it," or "weigh alternatives." Most teachers choose to write essay questions because they believe that they capture types of learning that objective-type questions miss. In Chapter 5 we learned that an essay exam includes several questions that require you to draw upon past learning to demonstrate "higher-order thinking." Other features of the essay exam are that you have a limited amount of time to answer and the outcome of your efforts is in written form.

Most measurement textbooks state that the principal disadvantage of the essay is scoring reliability. Studies comparing how experts score the same essay sometimes show poor agreement between or among the raters (Thorndike et al, 1991). More recently, however, many educators have concluded that the principal problem with the traditional essay test is construct relevancy (Alleman & Brophy, 1997).

We discussed construct relevancy in Chapter 4. As you recall, it deals with whether the test assesses what you want it to assess. As we pointed out above, most teachers develop essay tests to assess higher order thinking skills or deep understanding. But, when we closely examine these tests and the
context in which they are used, it appears that they are measuring something different (Baker, 1994; Cizek, 1997; Newman, 1997).

For example, let's say that you want to determine a student's ability to learn and understand new information at the end of a two week unit on rocks. You write the following essay question:

Identify the principal types of rocks and compare and contrast the forces that produced them. Be sure to discuss how you would determine what type of rock an unfamiliar rock is.

This task presents some serious obstacles to the valid assessment of "ability to learn and understand new information." First, knowledge and understanding of new content and writing skill will be difficult to separate. A learner could have met the goals of the rock unit and yet be unable to express this learning satisfactorily in writing, given the time and space constraints of an essay exam. But a greater challenge to construct relevancy is that no new information was given to the learner to learn and understand! The task depends completely on old information and is almost completely dependent on memory. Added to this, the student did tasks of a similar nature in class. Thus, based on this question, the teacher has no way of judging the learners ability to process new information.

We can't measure a learner's ability to acquire, organize and make sense of new knowledge (i.e., construct new knowledge) and to apply this knowledge to solve novel problems when the task depends almost exclusively on the recall of old knowledge. Thus, the restricted response essay exam with its time constraints, with the restrictions it places on answers, with its reliance almost exclusively on writing skill, and with its dependence on old
information often is unsuited to the assessment of *deep* understanding. A different assessment task is required.

Performance Assessment: Are These Activities Or Tests?

Go back and look over the sample tasks that began this chapter. Are these the activities used to teach deep understanding or the test questions used to assess it? A more important question is: Should there have to be a difference?

The traditional essay test takes place after a lesson or unit of study is completed. This practice, as we pointed out in Chapter 1, leads to some unfortunate results. Learners miss the connection between what happens in class and what happens on test day, reducing the motivation both to learn in class and to study for tests. In addition learners often do not receive feedback on the adequacy of their answers until sometimes long after their actual performance.

One of the key features of performance assessment, therefore, is that it can serve as a classroom activity or a test. Let's examine some other features of this type of assessment.

Performance Tests Are Direct Measures Of Learning

With most educational and psychological tests, the behavior that you see is not what you want to measure. The behavior that you see on a multiple-choice test in history is a learner drawing a circle around the letter that represents the best answer. Depending on how many correct choices the learner makes, the teacher judges the degree to which learning has or has not taken place. In other words, learning is measured indirectly. It is inferred from observing the learners test taking behavior.

But, with performance tests the behavior that we see is the learning that
we want to assess. A diving or gymnastics competition is a type of performance test. The behavior that the judges see is the same as the behavior that they want to assess. A performance test asks the learners to do the things you want to measure. You set up situations that allow you to observe and assess the learning of a problem-solving strategy directly. Or, you ask the learner to do certain things that allow you directly to observe and assess the learning of a scientific procedure. With performance assessment you don't infer that the learner can analyze, compare, plan, or evaluate. Rather, your test asks the learner to do these things and you observe and make judgments about the quality of the performance that you see.

Performance Tests Assess Processes And Products

When we create a performance test we are interested in assessing what a learner can do or accomplish. But we also want to assess how they accomplished it. We want learners to demonstrate writing competence so we ask them to produce a story or essay of some kind. But we also want to assess the process of writing—the planning, translating, and revising that goes into a good writing product. Likewise, in science we want to see that the learner can produce a science project and also use good scientific thinking to do it.

Thus, the typical performance test assesses the products of learning, a report, project, painting, film strip, etc. and the processes that produced them. It assesses knowledge use, skill use, strategy use and the products that result from these processes. In judging the quality of these two types of learning, you will make use of the checklists that you studied about in Chapter 6 and the rating scales discussed in Chapter 7.
Performance Tests Can Be Embedded in Lessons

Performance testing can occur during an examination period at the end of a term or semester. But, ideally, it should be integrated with classroom lessons or activities. In fact, as we indicated above, the ideal performance test is a good teaching activity (Shavelson & Baxter, 1992). For example, Figure 8.1, shows a performance activity and assessment that is part of a unit on electricity (Shavelson & Baxter, 1992). The students are learning important principles about electricity while, at the same time, the teacher observes and rates the learners on the methods they use to solve the problems. As you can see this type of assessment allows the teacher to give immediate feedback to learners. It also makes explicit to learners the link between teaching and testing.

Insert Figure 8.1 about here

Performance Tests Can Assess Social Skills

Although performance tests are ideally suited to assess knowledge, deep understanding, and problem-solving strategies, they can also be used to assess a learners work habits and social skills such as cooperation, sharing and negotiation. Many performance assessments are group projects where learners have different but interdependent roles and tasks to fulfill. Since the ability to get along and work with others is important for success as a journalist, weather forecaster, historian, mechanic, or park ranger, authentic performance tests are ideal opportunities to teach and assess these competencies. Chapter 10 will focus on the assessment of social skills.
What Is Performance Assessment?

Performance assessments have been called *authentic assessments*, alternative assessments, and expanded assessments. Basically, they are attempts to build tests that assess more than just knowledge, more than just simple understanding, more than just the automatic use of a procedure, and more than just the application of a problem solving strategy. They are tests that try to determine if a learner "really knows" about something, a quality which we have termed "deep understanding." They do this by challenging learners with tasks that ask them not simply to recall knowledge but to construct or organize it, not just to solve problems but to demonstrate a disciplined approach to doing something that requires strategic thinking and metacognition. And, they do this by embedding the assessment task in a context that has some meaning or purpose beyond school or beyond the bounds of the immediate classroom lesson or unit.

DEVELOPING A PERFORMANCE ASSESSMENT

When you decide to design a performance assessment you are, in a sense, committing yourself to certain theories of learning and instruction. Performance assessment is your way of validating the assumptions underlying these theories. These assumptions include the notions that students learn best in contexts where they struggle with relevant and meaningful tasks under the guidance of an instructor and often with the support of other learners. Included in these assumptions is the belief that the mind spontaneously stores and organizes information, and automates skills and strategies. In other words, you believe that learners construct their own declarative and procedural knowledge bases. Instruction can inhibit or promote knowledge construction but not prevent it. Obviously you want to facilitate it. And,
performance assessment is one way of finding out if you are doing that. Let's examine some general principles for designing performance assessments.

Performance Assessments Should Require Knowledge Construction

Performance assessments that are embedded in lessons accomplish two purposes: by being an integral part of the learning process they help learners acquire and construct new knowledge. In addition, they are a means of assessing cognitive activity. Performance tests force learners to approach a learning task with the goal of in-depth understanding. They have to know more than a lot of details about a topic: they have to see relationships and organize their knowledge bases accordingly.

When we construct knowledge, we organize it in ways that we talked about in Chapter 5. We detect or identify relationships among bits of information or concepts. We make comparisons, draw inferences or implications, and evaluate what we read or hear. We also extract generalizations, rules, or principles that aren't explicitly pointed out in the oral or written text. The only way to assess learners' ability to construct knowledge during a performance test is to give them new knowledge and ask them to do certain things that reflect this capability. Notice how knowledge construction uses the general cognitive strategies of analysis, comparison, inference and interpretation, and evaluation that you studied in Chapter 7.

Thus, an essential feature of any performance assessment will be a task that requires learners to acquire and then construct new information either by reading a speech or magazine article, examining a graph or chart, viewing a film, listening to a debate, or searching through an encyclopedia. If the test does not present new information, but requires solely the recall of prior information, then knowledge construction and deep understanding cannot be
validly judged.

Performance Assessments Should Require Strategic Thinking

When a historian is given some original historical document to interpret, or an historical analysis to critique, they engage in what is called strategic thinking or disciplined inquiry (Newman, 1997). In other words they draw upon their extensive prior knowledge and engage in certain problem solving processes (evaluate the credibility of the source, corroborate evidence, and contextualize interpretations). The same is true of mathematicians, journalists, and geographers.

Thus, a performance assessment that is valid for assessing deep understanding, requires learners to draw upon their prior knowledge base (both declarative and procedural) and use both general and specific problem solving strategies. It must explicitly cue students that they are to demonstrate their strategic knowledge. You practiced how to do this in Chapter 7.

Performance Assessments Should Require Clear Communication

The end result of a performance assessment should be a product of some kind. This product may be a written document, a map, a diagram, a demonstration that includes visual, written, and spoken material, or an oral presentation. In other words something tangible that enables the teacher to judge knowledge construction, strategic thinking, and skill in communicating ideas must be produced.
Performance Assessment Should Strive For Authenticity

As we explained in Chapter 1, authenticity has numerous meanings. One meaning is that the assessment task must relate to what real people do outside of the classroom. In other words, the math performance test must ask the learners to do something that mathematicians do as part of their job. The biology performance test must be something that professional biologists routinely do.

But we feel that this definition is unduly restrictive. The authentic task that you use in your performance test should be something that has value to the student beyond simply getting a grade that shows competence on a particular test. A task is authentic if it has job relevance. But, it also is authentic if it builds towards some higher learning that occurs later in the school year or in succeeding grades.

DESIGNING THE PERFORMANCE ASSESSMENT TASK

As with any test there are both conceptual and technical issues that must be resolved before developing a valid and reliable performance test. The conceptual issues concern the clear identification of the cognitive processes you want the test to measure. The technical issues concern task specification and scoring. Let's look closely at how these issues are resolved in designing a performance assessment.

Step One: Decide On A Specific Subject Area

The best way to get started is to pick a topic in your subject area for which you have deep understanding. In other words, choose a topic for which you have the critical declarative, procedural, and strategic knowledge necessary for deep understanding. If you are a biology teacher, you might
choose photosynthesis. If you are a fourth grade math teacher, you might choose fractions. The geography teacher might select how to locate a city. Or, the history teacher might assess deep understanding of the causes of the revolutionary war.

APPLICATION

In the space below, identify a topic that you feel sufficiently expert in to assess deep understanding of your learners:

________________________________________________________

________________________________________________________

________________________________________________________

________________________________________________________

________________________________________________________

Step Two: Define Those Cognitive Processes And Social Skills You Want To Assess

We develop performance assessments because we want to tap into a learner's higher order thinking or problem solving skills. At this stage you have to be clear about which higher order thinking or problem solving skills you wish to assess. Typically, performance tests assess cognitive processes involved in the acquisition and organization of information, the use of problem-solving strategies, and communication. For example, if I am
interested in assessing how well learners have acquired and organized information that they have just read or heard about, (i.e., knowledge construction) and communicate it clearly, I could ask them to do the following:

Read Senator Sharp's proposal on dealing with legal and illegal immigration. Identify the key principles underlying his proposal and compare them to those we've already studied in class. Write an essay explaining to a friend who is unfamiliar with the topic, the important concepts and facts about immigration, why immigration is a major issue in this country and how Sharp's proposal compares to others that you have studied.

This task requires students to extract the important elements of a proposal that they have never read before, make distinctions with what they already know, put the proposal in a historical context, and draw comparisons with other immigration policies.

In the area of mathematics, performance assessments can be designed to assess such cognitive processes as:

- Understanding and representing problems
- Discovering mathematical relations
- Organizing information
- Discovering strategies
- Formulating conjectures
- Using strategies
- Evaluating the reasonableness of answers
Here's an example of a mathematics performance test:

The following table shows the cost for different railroad fares:

<table>
<thead>
<tr>
<th>B&amp;O Railroad Fares</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>One Way</td>
<td>$2.00</td>
</tr>
<tr>
<td>Weekly Pass</td>
<td>$18.00</td>
</tr>
</tbody>
</table>

Sheryl is trying to decide if she should buy a weekly train pass. Three days a week, on Monday, Wednesday and Thursday, she takes the train to and from work. On the other two days she takes the train to work but gets a ride home with her sister. Should Sheryl buy a weekly train pass? Justify your answer.

This problem assesses understanding and representing problems, evaluating the reasonableness of an answer, and clear communication.

Here is an example of a performance test that assesses knowledge acquisition and organization, strategic thinking (analysis and comparison), and clear communication.

On the table in front of you are rocks of shale, pumice, marble, limestone, slate, sediment, granite, lava, and magma. Design a chart that identifies the different types of rocks, the classes they belong to, and that shows the relationships among the rocks. Draw arrows to
show the relationships and be sure to point out what each of these relationships mean and why they are important.

Performance tests often are implemented in a group learning format. In such cases, you want the learners to demonstrate collaborative or cooperative social skills. These can involve group members assigning and carrying out different roles as researcher, checker, recorder, trouble shooter, runner, and summarizer (Johnson & Johnson, 1991). If you choose to use a group context for your performance assessment, then, you should specify the collaborative skills that you want your learners to exhibit in the process of completing the performance task. Examples of such skills are:

- Praises performance of peers
- Summarizes what others have said
- Gives assistance when asked
- Shares materials
- Asks for feedback
- Paraphrases what others say

If we were to build a performance assessment to assess deep understanding of a 5th grade science lesson dealing with static electricity we would be interested in assessing understanding of the principles of attraction and repulsion, the ability to extract a principle concerning the movement of electrons, the cognitive strategies of planning, hypothesis testing, and prediction, skill in communicating the results of an experiment, and collaborative skills.
APPLICATION

In the space below describe the cognitive processes in the areas of knowledge construction, strategic thinking, clear communication and social skills that you are interested in assessing.

Step Three: Design The Task And Task Context

Having selected a topic area and identified the cognitive outcomes that you wish to assess, your next step is to create an authentic task, simulation, or situation for learners to show what they can do. You have seen several of these situations earlier in this chapter. Ideas for them come from newspapers, news programs, magazines, your experiences with a variety of job situations, or magazines and journals written for teachers on alternative teaching and testing formats.

They may involve presentations to a school board or city council, a reenactment of an historical event, a laboratory task, a science experiment, development of a travel brochure or a mass-transit plan. Many teachers get
started on developing the assessment context by asking themselves the following questions (Wiggins, 1992):

What do the jobs of professionals who make their living as mathematicians, writers, geographers, artists, or historians look and feel like?

Which of the projects and tasks of these professionals can be adapted for school instruction?

What skills do these professionals acquire on-the-job that learners can master in the classroom?

Once you start asking and answering these questions a host of ideas arise. The problem then becomes one of developing the task specifications. Here are some criteria that you should keep in mind when planning and implementing performance assessment:

**Goal relevance.** Will learners see the assessment as an integral part of classroom instruction or something that gets them a grade and little else? Do the assessment tasks reflect what you and your learners value? Do they ask for skills that learners have learned about and practiced in their lessons? This is probably the single most important design consideration.

**Level of difficulty.** Just as the track coach asks his sprinters to run 100 and 200 meter dashes in competition and not 800 meter runs, so, too, teachers must aim their assessment tasks within the optimal range of learner
competence. The tasks must be novel but familiar, representing something that was practiced but not identical to it. The activities should be challenging but with content with which they are accustomed. The purpose of the assessment is to have learners demonstrate the higher-level thinking skills that they learned during instruction. If the procedures they use and the content are too difficult, your learners will not be able to show what they can do.

Multiple goals. Real world tasks require a variety of thinking skills: analysis, interpretation, evaluation, planning, revision, clear communication, and self-monitoring. The same is true for a performance assessment. Although these assessments can measure one of the above skills, a focus on multiple goals is preferable. Learners find complex activities more engaging. Furthermore, performance assessments take time to develop. Teachers use them more efficiently when they direct them to accomplish many goals rather than one.

Performance assessments are ideally suited to incorporate content and skills from a variety of academic disciplines to accomplish these goals. They can require learners to integrate math, writing, science, and art skills to produce a project or demonstration.

Multiple solutions. In a performance assessment, we want learners to wrestle with a complex problem that can be solved in any number of ways. They must make and defend their choices of strategies to solve problems and methods used to communicate their solution. Learners should be able to select the data sources they consult (experts, magazines, encyclopedias, newspapers, scientific journals), and presentation formats, such as videotape, audiotape, persuasive essay, oral explanation, graphic displays, stories, or
dialogues. A key goal should be to adapt the assessment to the strengths and needs of the learner.

**Self-determined learning.** Deci et al. (1991) believes that one of the most important outcomes of classroom lessons should be feelings and beliefs about self-determination. Learners feel self-determination when they are allowed to master and demonstrate competence with a complex task, depending largely on their own intellectual resources and some assistance from peers and teachers. Performance assessments should be designed with self-determination in mind. In other words they should allow for a considerable degree of learner autonomy (i.e. choices), provide sufficient time and resources to allow learners to show their competence, and permit consultation with peers and adults.

**Clear directions.** Performance tests should be complex, require higher-level thinking, assess multiple goals, and permit considerable latitude about how to reach these goals. Nevertheless, your directions and descriptions of the project should leave no doubt in the minds of learners about what is expected. While your students need to think long and hard about how to carry out the task, they should be clear about what a finished product looks like. In other words, they should be able to explain exactly what you expect them to turn in when the assessment is over.

Here are the task specifications for our static electricity performance assessment:

The material in this box can be used to construct an electroscope similar to the one you have used in our class lessons. Construct an
electroscope with your group. It can be built in a lot of different ways. Assign roles to different group members for carrying out this project. Once you have done this, use the plastic ruler and the wool cloth to solve the following problems:

(1) What will happen to the leaves of the electroscope as the charged ruler is brought closer or farther away from it? Explain why this happens.

(2) What will happen to the leaves when you touch the electroscope with the charged ruler and then touch it with your finger? Explain why this happens?

(3) What will happen when you pass the charged ruler close to the leaves but do not make contact with them? Explain why this happens?

(4) What will happen to the charged leaves when you heat the air next to them with a match? Explain why this happens?

For each question, fill out the appropriate section on your lab sheet:

(a) Make a prediction about what will happen?

(b) What did you observe?

(c) Was your prediction supported?

(d) Explain what you saw using the important terms that we
studied in class?

(e) Make a drawing that shows what you did and the electrical forces at work?

Now you be the judge of whether this task meets the following criteria of task design that we just discussed: goal relevance, level of difficulty, multiple goals, multiple solutions, self-determined learning, and clear directions.

APPLICATION

In the space below, identify the task and task context for the performance assessment that you are developing:

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

Step Four: Specify The Scoring Rubrics

Performance tests like the ones above can be scored reliably, but take time to do so. Unlike the objective-type test questions that you learned about
in Chapter 5, there is no right or wrong answer to a performance test. You must make difficult decisions about the characteristics of the processes and products that you want to assess and the relative importance of each of these characteristics in your overall evaluation of learner performance.

Given these obstacles to reliable scoring, it is tempting to limit your scoring criteria to those aspects of performance that are easiest to rate (e.g. Did the learner make a drawing? Did the learner complete the lab sheet? Did the learner complete the project on time?). Such "yes-no" judgments ignore the more important characteristics that you identified in Step 2: clear communication, knowledge-construction, strategic thinking, planning and revision, etc. By focusing on these processes you guard against turning an authentic performance test into a knowledge test. Thus, your goal when designing a scoring system (i.e., the scoring rubrics) is to do justice to the effort that you expended in developing the performance assessment and that your learners exerted when taking it.

By giving careful consideration to this step of the performance test development process, you will minimize scoring subjectivity and bias while, at the same time, holding learners to high standards of achievement. Let's examine some of the important considerations in developing rubrics for a performance test.

**Measure your goals.** In Step 2 you identified the learning outcomes that you want to assess with the performance test. Your rubrics should fit these accomplishments. Will you be expecting evidence of knowledge acquisition and organization? If so, you will need to design a rating scale that reflects this. Do you wish to assess the use of problem solving strategies? If so, you will need rating scales to assess their use. Will you be observing the
processes that learners employ as they complete the assessment. If so, you may need to construct checklists that record whether the processes were used. Do you also expect a product of some kind? If so, you will need to construct rating scales that assesses the qualities that you desire in the poem, written essay, exhibit, map, graph, etc. to be produced. In Chapters 6 and 7 you learned how to develop a variety of scales and checklists to make assessments such as these.

**Select an appropriate scoring system.** Choose a scoring system that is suited for the type of outcome that you want to measure. As you have learned, there are three categories of rubrics to consider: checklists, rating scales, and holistic scoring, as shown in Figure 8.2. Since performance tests measure multiple goals, you may need all three types. Let's review their distinctive features.

---

Insert Figure 8.2 about here

---

**Checklists.** Checklists contain lists of behaviors, traits, or characteristics that can be scored as present or absent. They are best suited for complex behaviors or performances that can be divided into a series of clearly defined specific actions. We gave examples of these in Chapter 6. They take time to develop but have good scoring reliability.

**Rating scales.** Rating scales are typically used for assessing complex products and some processes that do not lend themselves to yes/no or present/absent type of judgments. As you learned in Chapter 7, the most common type of rating scale assigns numbers to certain levels or degrees of
performance. These scales focus the rater's attention on certain aspects of the product or process to which the rater assigns a number that represents important qualities of learning.

The key question that you ask when designing a rating scale is: What are the most important characteristics that show a high degree of the thinking process, learning, or trait being measured? In other words, what are the important aspects of clear communication, knowledge construction, problem solving, or metacognition that I want to assess? You had practice at specifying these characteristics in Chapter 7.

*Holistic scoring.* Holistic scoring is used when the rater estimates the overall quality of the product, demonstration, or exhibit and assigns a numerical value to that quality, instead of rating the various characteristics of the performance. Holistic scoring is typically used in evaluating extended written products like stories or term papers, or artistic performances like a dance or reading interpretation. You may decide to score your class' term papers on an A to F scale. In such a case, it is important that you have model papers that exhibit the varying degrees of such a scale. After having created or selected these models from the set to be scored, you again read each paper and assign each of them to one of the categories. A model paper for each category (A-F) helps to assure that all the papers assigned to a given category are of comparable quality.

Holistic scoring can be more difficult to use for processes than products. For the former, extensive experience in rating the process or performance is necessary (e.g., rating a dramatic rendition of a poem, oral interpretation of a story, or debate). In such cases, you may find it helpful to make audio or videotapes of the performance for other classes and school years as models representing different degrees of competence. Another
characteristic of holistic scoring is that it is difficult to give the learner detailed feedback on the quality of their performance.

As we stated in Step 3, good performance tests require learners to achieve multiple goals in alternative ways. Thus, a comprehensive performance assessment typically combines all three rubric scoring systems.

**Assigning point values.** You will have to decide how many points to assign the project as a whole, and how many points to assign to each component of the project (communication, understanding, planning, presentation, etc.). We will discuss how to decide the total value of the project to your overall grading system in Chapter 11 when we discuss classroom grading systems.

It is a good idea to limit the number of points each component is worth to that which can be reliably discriminated. For example, if you assign 20 points to how well the person communicates the results of the project, you are implying that you can reliably grade 20 degrees of communication quality. Few raters can discriminate 10 degrees of quality on any trait, much less 20.

You have two related goals in assigning points: Assign enough points to a component or trait so that you do justice to varying degrees of quality; but avoid assigning so many points that the rater starts assigning points arbitrarily. We recommend that you assign no fewer than 3 and no more than 7 points to any one component of a performance assessment.

**SCORING RUBRIC FOR STATIC ELECTRIC PERFORMANCE TEST**

**Checklist for social skills (Assign 1 or 0 points)**

8-25
___ Role differentiation was accomplished
___ Gave praise to a peer
___ Carried out role
___ Asked for clarification
___ Used expressions like, "What I meant to say was..." or "Do you agree with what I suggested?"

Total points __________

Rating Scales for Cognitive Processes

Knowledge of Basic Concepts of Static Electricity

(Assign 1-5 points)

<table>
<thead>
<tr>
<th>Scale</th>
<th>Description</th>
</tr>
</thead>
</table>
| 5     | Firm command of basic concepts  
        Uses terminology correctly  
        Identifies important principles |
| 4     | Shows nearly complete understanding of basics.  
        Most terms used correctly  
        Has identified most important principles |
| 3     | Has only tentative grasp of concepts  
        Some terms used incorrectly  
        Some inference evident. |
| 2     | Lacks command of most of the important concepts  
        Uses little relevant terminology  
        Little evidence of ability to abstract principles |
| 1     | Shows no understanding of basic concepts  
        No attempt to use relevant terminology  
        No evidence of abstraction or inference |

8-26
Points ____

Carrying out the task  
(Assign 1-5 points)

<table>
<thead>
<tr>
<th>Scale</th>
<th>Description</th>
</tr>
</thead>
</table>
| 5     | Very carefully planned project  
        Stated clear hypotheses  
        All predictions are explicitly stated |
| 4     | Thorough planning and anticipation of most problems  
        Hypotheses are for the most part clear  
        Most predictions are clear and explicit |
| 3     | Evidence of some planning  
        Hypotheses stated but vague  
        Attempted to make clear predictions |
| 2     | Little planning went into this  
        Hypotheses incomplete and unclear  
        Some vague predictions made |
| 1     | No evidence of planning  
        Does not make hypotheses  
        No predicted results |

Points ____

**Rating scale for communication**

<table>
<thead>
<tr>
<th>Scale</th>
<th>Description</th>
</tr>
</thead>
</table>
| 5     | Gives a complete, clear, unambiguous report  
        Includes diagrams and other visual aids.  
        Uses good organization and mechanics. |
| 4     | Gives a fairly complete report that is generally clear |

8-27
Uses some visual aids.
Organization and mechanics are generally good.

3
Report is somewhat ambiguous and not always clear
Could use more visuals that are clear.
Presentation is a little disorganized and writing weak.

2
Report is difficult to follow.
Very limited visuals; hard to understand
Poor organization with noticeably poor
writing mechanics.

1
Ineffective communication. Incomprehensible.
No visuals of any kind
No evidence of any organization; numerous
spelling and grammatical errors.

Points ______

Total Points ______ (Maximum of 20)

APPLICATION

In the space below, describe a set of scoring rubrics for your performance assessment.

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

8-28
Step Five: Identify Important Implementation Considerations.

In carrying out Steps 1-4, you have created the performance task. Now it's time to think through how you will implement it. There are two major considerations at this time: (1) specifying testing constraints; and (2) planning how you will deliver the assessment activities to the learner.

**Identifying testing constraints.** Testing constraints refer to the conditions under which learners demonstrate their deep understanding. Here are some of the most common forms of testing constraints:

*Time.* How much time should a learner or group of learners have to plan, revise, and finish the task?

*Reference Material.* What resources will learners be able to consult while they are completing the assessment task? Dictionaries, textbooks, class notes, CD-Roms?

*Other people.* Will your learners be able to ask for help from peers, teachers, and experts as they take a test or complete a project?

*Equipment.* Will your learners have access to computers, calculators, spell checkers, etc. as they complete the assignment?

*Scoring criteria.* Will you inform your learners about the explicit standards that you use to evaluate the assessment?

In deciding which of these constraints to impose during a performance
assessment, ask yourself the following questions:

What constraints do people in the real world face when they do these type of tasks?

What kinds of constraints tend to elicit the best performance in apprentice performers?

What are the appropriate limits a teacher should impose on the availability of the five resources listed above?

APPLICATION

In the space below, identify the testing constraints that will be imposed during your performance assessment.

__________________________

__________________________

__________________________

__________________________

Delivering the assessment. There are five important events that need to occur for successful implementation of a performance assessment. They are structuring the task, motivational set, initial coaching, independent work, and
debriefing.

Structuring. How will you describe the performance task to your learners? What words and expressions will you use to communicate the goals and purposes of the activity? What will you say to get them to recall relevant prior knowledge and cognitive strategies? What will you do to make sure that they understand the task requirements before they begin the assessment: what they will be doing, what product they will produce, and how their performance will be evaluated?

Motivational set. The key to the effectiveness of the performance assessment is learner engagement, not simply performing physical actions or time on task. What will you say and do to stimulate their curiosity and interest and help them to see the importance of an all-out effort?

Coaching. What explanations and modeling will you give before setting learners out to work on their own or in small groups? Will you have to guide them through several examples and work through several more examples before they work independently?

Independent work. What monitoring will you do once learners are working on their own? Will you be giving them any type of formative feedback while they are completing the assessment task? How will you handle confusions or misconceptions that you notice while they are completing the task?

Debriefing. How will learners receive feedback on their performance? Will they be able to learn from their mistakes? How will you tie their performance to the goals and purposes of other classroom activities? What will you do to help learners reflect on what they have learned and accomplished? How will you help them interpret their accomplishments in
light of the big picture beyond the immediate task requirements? In other words, how will you help them link the performance assessment with the larger unit or curriculum strand, later learning, or the world outside the classroom?

APPLICATION

In the space below, describe some ideas on how you plan to address the implementation issues of structuring, motivation, coaching, independent work, and debriefing?

Structuring:

________________________________________________________


Motivation:

________________________________________________________


Coaching:

8-32
Independent work:

Debriefing:

Final Comments

A well designed and implemented performance assessment will tell you things about your students' understanding and achievement that no other type of assessment can rival. But, even more importantly, it can represent a type of motivating exercise that reflects the permanency and synthesis of prior learning. When performance assessments are authentic, they become as highly motivating to learners as the final track championships or band competitions are for their contestants.

In the next chapter you will learn about another type of performance assessment that can serve a similar motivating function. It is called *portfolio*
assessment and is uniquely designed to assess not only deep understanding and synthesis, but also changes in learner performance across the school year.

Activities
1. If you were to test your learners for domain-general problem solving strategies, what rules would you use to guide the development of your assessment instrument?

2. In earlier chapters you have studied lists of behaviors provided by Bloom et al. and Gagne, some of which purport to measure higher-order problem solving behavior. If you were to build a taxonomy of your own, not using any terms from Bloom et al. and Gagne, what behavioral descriptions might your taxonomy of problem solving behavior include and what would be the hierarchy of behaviors, if any?

3. Assume you are planning a lesson to teach metacognitive strategies using the model provided by Quelmalz. Describe what your lesson would do to get learners to plan (analyze the problem), draft and tryout (attempt to solve the problem), monitor and revise (check whether goals are achieved and the strategies working), and evaluate and reflect (check the adequacy of the solution and effectiveness of the strategies used).

4. Assume you want to teach your learners to use four steps to analyzing a poem: (1) identify key words, (2) look up in a dictionary the meaning of key words, (3) paraphrase what you think the author is saying, and (4) relate what the author is saying to your own experience. Develop an assessment instrument with definitions of the behaviors you want to measure that assesses the extent to which learners have applied each of the four steps.

5. Choose a topic for a lesson you will be teaching. Identify all the behaviors
a worker in the real world is likely to use on the job related to this topic. Which of these could be addressed within the goals of your lesson?

Suggested Readings

Excellent primer on performance assessment. Explains clearly the different types of performance tests and how they can be used to evaluate instruction.
Chapter 9
Assessing Genuine Achievement: The Portfolio

We have woven a number of themes throughout the fabric of each chapter of this book. Some of these themes have been:

- Assessment must align with instruction.
- The skills that learners demonstrate on tests should be the skills that are practiced in class.
- Motivation for learning and doing well on tests spring from how the assessment program is designed.
- Students become more engaged with teaching and testing when assessment is authentic.
- Learners like to grapple with challenging tasks.
- A good assessment measures not only the products of learning but also the processes that generate them.

As we saw in the last chapter, performance assessment is an exercise that exemplifies all of these themes. As you will recall, performance assessment is a type of demonstration by which learners show their deep understanding of a particular area of learning. This demonstration is like a snapshot that captures what a learner has accomplished at a particular point in the academic year.

There is another type of performance assessment that is more than a one time picture of what a learner has accomplished. It's principal purpose is to tell a story of a learner's growth in proficiency, long-term achievement, and significant accomplishments in a given academic area. It is called portfolio
assessment. The portfolio is a measure of deep understanding like the performance demonstrations covered in Chapter 8. But, in addition, it shows growth in competence and understanding across the term or school year.

Portfolio assessment is based on the idea that a collection of a learner's work throughout the year is one of the best ways to show both final achievement and the effort put into getting there. You are already familiar with the idea of a portfolio. Painters, fashion designers, artisans, and writers assemble portfolios that embody their best work. Television and radio announcers compile video- and audio-taped excerpts of their best performances that are presented when interviewing for a job. A portfolio is their way of showing what they can really do.

Classroom portfolios serve a similar purpose. They show-off a learner's best writing, art work, science projects, historical thinking, mathematical achievement. But, they also show the steps the learner took to get there. They compile the learner's best work. But they also include the works-in-progress: the early drafts, test-runs, pilot studies, or preliminary trials. Thus, they are an ideal way to assess final mastery, effort, reflection, and growth in learning that tell the learner's "story" of achievement.

The idea of classroom portfolio assessment has gained considerable support and momentum. Many school districts use portfolios and other types of exhibitions to help make decisions about promotion and graduation. These are examples of using portfolios for what are termed "large-scale" or "high-stakes" assessment (Stecher, and Herman, 1997). Research points out that portfolios have potential for supplementing the information that standardized tests provide about learners. Nevertheless, many researchers express reservations about the reliability of portfolios for making important decisions about learners related to graduation or college admission.
But, the purpose of classroom assessment is not to make high-stakes decisions about learners. Rather, it is to motivate effort and show achievement and growth in learning. While the reliability and validity of a classroom teacher's judgments are always a matter of concern, they are less so when the teacher has multiple opportunities to interact with learners, and numerous occasions to observe their work and confirm judgments about their capabilities.

Thus, the purpose of this chapter is to demonstrate the potential of portfolio assessment for accomplishing the learning goals that you have for your students. We will first clarify what portfolio assessment is. Then, we will cover the most significant design considerations: deciding on the purpose of the portfolio, the cognitive outcomes to be assessed, who will plan it, what products to include, the criteria for assessing outcomes, the data that need to be collected to document progress, effort, and achievement, the logistics of where the products are kept, and finally, how collaborative feedback will be given.

ACT I, SCENE 1: PLANNING FOR SOMETHING NEW

Keith Jefferson is about to start his seventh year of teaching eighth grade honors English. It is August and the last day of a two-week continuing education program for teachers on classroom portfolio's. The instructor, Maria Peron, is starting to wrap things up:

*Ms Peron:* So the important thing to remember is that you don't decide on Friday to do portfolio assessment and start it on Monday

As Keith leaves the seminar and enters his empty classroom, he is
apprehensive. He wants to do something different this year, but doubts that he can pull it off. He has a system for teaching writing that took six years to develop and is working well. "If it's not broke, don't fix it," he says to himself. But, as Maria said, and as many teachers like him have done, why not just get started and work things out as they go along--but with a good plan at the start.

ACT I, SCENE 2: FIRST DAY OF CLASS

During the first 15 minutes of class, Keith went over the general rules and routines that he expects his learners to follow. He's now explaining his goals and class requirements:

Keith: My goal this year is to help you be show-offs, to show you how to "strut your stuff." Ice skaters are show-offs...so are painters, and track stars, and guitar players. Only in this class you're going to show-off your writing. At the end of this class you're going to collect examples of your best stuff and show it off. But you're not showing it off for me. You'll be showing it off to someone else. It may be your parents, a magazine or newspaper editor, a writer you respect, a special high-school you want to get into, an employer that you want to hire you, or may be a television producer for whom you want to develop a new program. But, the most important purpose of this class is to collect a sample of your best writing that shows what you can do when you really try. We'll call this collection of your work your portfolio. Your portfolio will be your story, your one-person-show, your exhibition of your best stuff.

Joan: But how do we know what goes in it?
Keith: That's mostly up to you to decide. I'll give you some basic requirements but it's largely up to you. I think it would be a good idea if people see not only your best work, but how much work you put into doing it. So, you'll be putting first drafts and second drafts in your portfolio so that someone can see how much your writing improved?

Tyrone: Like what kind of things do we put in it?

Keith: We'll be learning about a lot of different kinds of writing this year: poetry, essays, writing where you're trying to persuade someone, criticism, autobiography, short fiction, dialogue. You'll select among these types and decide which writing pieces you want to put in it.

Samantha: And, what kinds of grades do you give?

Keith: I want to know and your parents want to know how clearly you write, how well you plan, how flexible you are when revising. My grades will tell how well you do. I'll be explaining how I assign the grades as you begin to design the portfolio. Any other questions?

O.K. For the next twenty minutes we'll break into small groups and you'll discuss among yourselves what you want the portfolio to say about you, what you want in your portfolio, and who you want to see it. Each of you will put your personal choices on this form, bring it home tonight, show it to your parents, and discuss your portfolio with them. Return this on Thursday. I've also prepared a handout for you and your parents that explains exactly what a portfolio is and how you will build one.
Keith quickly organizes the groups, gives them some directions for discussion, and hands out a form to plan the portfolio.

This scenario illustrates the first steps in portfolio design: Deciding (1) the purpose for which the portfolio will be used; (2) what cognitive skills and dispositions the portfolio will assess; (3) who will plan the portfolio; and (4) what work will go in it. We will discuss each of these considerations. But, first, let's explore what we mean by the expression "using portfolio's for classroom instruction and assessment."

Rationale For The Portfolio

Many states and school districts are experimenting with portfolios as a tool for making decisions about promotion, graduation, and college admission--so called "high-stakes" decisions. We are proposing a different purpose for portfolios. We believe that their greatest potential is for showing teachers, parents, and learners a richer array of what students know and can do than paper-and-pencil tests and other "snap-shot" assessments.

If designed properly, portfolios can show a learner's ability to think and problem solve, to use strategies and procedural-type skills, and to construct knowledge. But, in addition, they also tell something about a learner's persistence, effort, willingness to change, skill in monitoring their own learning, and ability to be self-reflective or metacognitive. So, one purpose for a portfolio is to give a teacher information about a learner that no other measurement tool can provide.

But, there are other reason's for using portfolios. They can alter the nature of classroom instruction by changing the relationship between teacher and learner. As was evident in the classroom scenario, Keith was not just
making a commitment to teach writing. He was making a commitment to be a coach or mentor. Just as promising young musicians, ice skaters, swimmers, actors, or gymnasts seek to apprentice themselves to masters, likewise, Keith's learners will look to him to help tell their story. Keith is setting up his classroom to be a "cognitive apprenticeship" (Collins, Brown, & Newman, 1989).

Portfolios are also means to communicate to parents and other teachers the level of achievement that a learner has reached. Report card grades give us some idea of this. But, portfolios supplement grades by showing parents, teachers, and learners the supporting evidence.

Portfolio's are not an alternative to paper-and-pencil tests, essay tests, or performance tests. Each of these tools possesses validity for a purpose not served by a different tool. If you want to assess a learner's factual knowledge base (as discussed in Chapter 4), then objective-type tests are appropriate. If you are interested in a snapshot assessment of how well a learner uses a cognitive strategy, there are ways to do this that don't involve the work required for portfolio assessment. But, if you want to assess both achievement and growth in an authentic context, portfolios are a tool that you should consider.

Finally, portfolios are a way to motivate learners to higher levels of effort as we discussed in Chapter 2. They provide a seamless link between classroom teaching and assessment in a way that is consistent with modern cognitive theories of learning and instruction.

Definition of a Portfolio

Our definition of a portfolio is the following:
A portfolio is a planned collection of learner achievement that documents what a student has accomplished and the steps taken to get there. The collection represents a collaborative effort among teacher, learner, and parent to decide on portfolio purpose, content, and evaluation criteria.

Designing The Portfolio

As Maria Peron pointed out, portfolio assessment takes planning. Otherwise, it may end up being a drawer full of manila folders that students never give a second thought to until they get a grade. When properly designed, portfolios can accomplish all that we talked about above. When carelessly designed, they can be misleading to parents, teachers, and students as to what was accomplished. Before getting into portfolio design, let's review this issue as it relates to portfolios.

Insuring Validity of the Portfolio

Let's say that one of the goals you have for the portfolio is to assess how well learners can communicate to a variety of audiences. However, you collect only formal samples of writing of the type you would submit to a literary journal. Or, you want your math portfolio to assess growth in problem solving ability. Yet your evaluation criteria place too heavy an emphasis on the final solution. These are some of the pitfalls that can undermine the validity of the portfolio. In general, there are three challenges to validity that you need to address: representativeness, rubrics, and relevance.

**Representativeness.** The best way to insure representativeness is to be clear at the outset about the cognitive learning skills and dispositions that you
want to assess and to require a variety of products that reflect these. You want
to insure that the samples of writing, scientific thinking, mathematical
problem solving, or woodworking reflect the higher-order thinking skills,
procedural skills, or dispositions that you want the portfolio to measure.

**Rubrics.** You have already had practice at designing rubrics in the
previous chapter. The same considerations for designing clear criteria to
assess complex performances or demonstrations also apply to assessing
portfolios. You will want criteria for assessing both individual entries and the
portfolio as a whole. We will cover this issue more extensively below.

**Relevance.** Assembling the portfolio shouldn't demand abilities of the
learner extraneous to the ones you want to assess. A second grade geography
portfolio whose purpose is to reflect skill in map making shouldn't demand
fine motor skills beyond what you would expect a seven year old to possess.
Likewise, a junior high school science portfolio designed to reflect problem
solving shouldn't require the reading of scientific journals that are beyond the
ability of a ninth grader to understand. As emphasized in Chapter 4,
measurement devices often lack validity because they require learner skills
that are extraneous to those the instrument was built to measure.

Now that you've given some consideration to validity, let's get started
on building a system for portfolio assessment for your teaching area.

Step 1: Deciding on the Purposes for a Portfolio

In the scenario above, Keith wants his learners to think about their
purpose in assembling a portfolio. Having learners identify for themselves the
purpose of the portfolio is one way to increase the authenticity of the task.
And, we encourage you to use this as part of your teaching strategy. However,
your learners' purposes for the portfolio (e.g. getting a job with the local news
station), won't necessarily coincide with yours (e.g. evaluating your teaching). In this section we are concerned that you be clear about your purposes at the outset of portfolio design.

Classroom level purposes that portfolios can achieve include:

☐ Monitoring student progress
☐ Communicating what has been learned to parents
☐ Passing on information to subsequent teachers
☐ Evaluating how well something was taught
☐ Showing-off what has been accomplished
☐ Assigning a course grade

APPLICATION

In the space below, identify the purposes that you would want a portfolio in your grade or content area to achieve:

1. 

2. 

3. 

4. 

Step 2: Identifying Cognitive Skills and Dispositions

Portfolios, like performance assessments, are measures of deep
understanding and genuine achievement. They can measure growth and development of competence in areas like knowledge construction (e.g. knowledge organization), cognitive strategies (analysis, interpretation, planning, organizing, revising), procedural skills (clear communication, editing, drawing, speaking, building), metacognition (self-monitoring, self-reflection), as well as certain dispositions—or habits of mind—like flexibility, adaptability, acceptance of criticism, persistence, collaboration, desire for mastery. Throughout this text you have had practice in specifying different types of cognitive learnings, identifying aspects of these learning types, and in planning to assess them. Apply this same practice to specifying what you want to know about your learners from their portfolios. As part of your teaching strategy, you will want to discuss these outcomes with your learners.

APPLICATION

In the space below identify the cognitive learning outcomes (e.g. metacognitive skills), several important behaviors (e.g. self-reflection, planning), and significant dispositions (e.g. flexibility, persistence) that will be reflected in your learners' portfolios.

Outcomes: 

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

Behaviors: 

________________________________________________________________________
Step 3: Deciding Who Will Plan the Portfolio

The principal stakeholders in the use of the portfolio are you, your learners and their parents. Therefore, it makes sense that all have a role in planning. Keith handled this issue by having learners identify their goals for the portfolio. He involved parents by sending home an explanation of portfolio assessment. In addition, he asked that parents and students discuss its goals and content.

When deciding who will plan the portfolio, consider what's involved in preparing gymnasts or skaters for a major tournament. The parent hires a coach. The coach, pupil, and parent plan together the routines, costumes, practice times, music, and so on. They are a team whose sole purpose is to produce the best performance possible. The gymnast or skater wants to be the best that he or she can be. They also want to please their parents and coach and meet their expectations. The atmosphere is charged with excitement, dedication, and commitment to genuine effort.

This is the atmosphere you are trying to create when using portfolios. You, the learner, and parents are a team for helping the student improve writing, or math reasoning, or scientific thinking and to assemble examples of
this growing competence. Learners want to show what they can do and to verify the trust and confidence that you and their family have placed in them. The portfolio is their recital, their tournament, their competition.

APPLICATION

In the space below, identify who will plan the portfolio assessment that you are designing, and describe how you will involve each participant in the planning process.

Who: ____________________________________________

__________________________________________

__________________________________________

__________________________________________

How:

__________________________________________

__________________________________________

__________________________________________

__________________________________________

9-13
Step 4: Deciding Which Products to Put in the Portfolio and How Many Samples of Each Product.

There are two key decisions to be considered: ownership and your portfolio's link with instruction. Ownership refers to your learners' perception that the portfolio contains what they want it to. You have considered this issue in Step 3. By involving learners and their parents in the planning process, you enhance their sense of ownership. You also do this by giving them a say in what goes into the portfolio. The task is to balance your desire to enhance ownership with your responsibilities to see that the content of the portfolio measures the cognitive skills and dispositions that you identified in Step 3. This brings us to the second key consideration: your portfolio's link with instruction.

Both learners and their parents need to see that your class instruction focuses on teaching the skills necessary to fashion the portfolio's content. You don't want to require products in math, science, or social studies that you didn't prepare learners to create. If it's a writing portfolio, then your instructional goals must include teaching skills in writing poems, essays, editorials, or whatever your curriculum specifies. The same holds for science, math, geography, or history portfolios. Thus, in deciding what you would like to see included in your learners' portfolios, you will have to insure that you only require products that your learners were prepared to develop.

The most satisfactory way to satisfy learner needs for ownership and your needs to measure what you teach is to require certain categories of products that match your instructional purposes and cognitive outcomes, and
to allow learners and parents to choose the samples within each category. For example, you may require that an eighth-grade math portfolio contain the following categories of math content (Lane, 1993):

1) "Number and operation," in which the learner demonstrates the understanding of the relative magnitude of numbers, the effects of operations on numbers and the ability to perform those mathematical operations;

2) "Estimation," in which the learner demonstrates understanding of basic facts, place value and operations; mental computation; tolerance of error; and flexible use of strategies; and

3) "Predictions," in which learners demonstrate abilities to make predictions based on experimental probabilities; to systematically organize and describe data; to make conjectures based on data analyses; and to construct and interpret graphs, charts, and tables.

Learners and their parents would have a choice of which assignments to include in each of the categories listed above. For each sample the learner includes a brief statement about what it says about his or her development of mathematical thinking skills.

Another example could be a high-school writing portfolio. The teacher requires that the following categories of writing be in the portfolio: persuasive editorial, persuasive essay, narrative story, autobiography, and dialogue. Learners choose the samples of writing in each category. For each sample they include a cover letter that explains why this sample was chosen and what they show about the learner's development as a writer.

You will also have to decide how many samples of each content
category to include in the portfolio. For example, do you require two samples of persuasive writing, one of criticism, three of dialogue, etc.?
Shavelson, Gao, and Baxter (1991) suggest that at least eight products or tasks over different topic areas may be needed to obtain a reliable estimate of performance from portfolios. It would be best for you to make this decision. Otherwise, different students and their parents might require different numbers.

APPLICATION

In the space below, identify in what general curricular area you will plan your portfolio (science, geography, reading, math) and describe how you will make decisions about which content areas and how many samples within each area to include. Be sure to include several categories of content from which learners (and parents) will choose representative samples.

Curricular area ____________________________________________________________

______________________________________________________________


Content areas __________________________________________________________

______________________________________________________________


9-16
ACT TWO, SCENE 1: THIRD WEEK OF CLASS

Over the weekend Keith reviewed the portfolio plans of each of his learners. He was pleased with the results. He had the learners with their parents choose five categories of writing from the seven that would be the focus of class instruction. He also listed on the form the cognitive skills and dispositions he hoped to develop with the use of portfolios and asked that they rate how important each one was for their development as a writer. He especially liked the reasons given for their choices of categories. Over the past two weeks, his lessons centered around the qualities of different types of writing.

He returned the forms to each student at the start of class along with some comments, and asked them to review the forms as he took attendance. Following some brief announcements over the loudspeaker, he began the days lesson:

Keith: On the forms I passed back I asked you to indicate what you want in your portfolio and what you want the portfolios to tell about you as a writer. Joni, share with us some of the things you and your parents said:
Joni: We'll I think it would be cool to work for a newspaper so I want some persuasive writing in my portfolio and good descriptive writing. I want the portfolio to show how hard I work and that I can make things interesting to read.

Keith asked several other students to describe their portfolios and then explained the day's lesson.

Keith: For the past three weeks we've been talking about what makes a good poem, short story, and so on. Now look at your notes and tell me what we said about poetry. Tobie?

Tobie: Good word choice, imagery, metaphors, and a central idea.

Keith: Good. Today, I want you to get into your groups, take the standards for different types of writing, and describe what a good example of writing, an average example, and a really poor one would look like.

Keith gave some examples of how he wanted his students to construct their rating scales and started them on the task. It took several days to do this. But, by the end of the week each student constructed his or her own summary rating sheet for each category of writing in the portfolio. Keith's plan was to have them take the form home, get their parents' input, and return them. He planned to review and edit them with each student and then make a final form. Here's an example of one of them.

Step 5: Building the Portfolio Rubrics.

In Step 2 you identified the major cognitive skills and dispositions that your portfolio will measure. In Step 4 you specified the content categories that your portfolio will contain. Now, you must decide what good, average,
and poor performance look like for each entry in the portfolio and the portfolio as a whole. Keith has decided to involve his learners and their parents in criteria development. He will take final responsibility for developing the criteria, revising, adding, or deleting student and parent criteria to best meet the goals of his writing class.

You already have experience at this. In Chapters 7 and 8 you identified the important characteristics of problem solving strategies and performance assessments that you wanted to teach and measure. You will follow the same process here. For each cognitive learning outcome for each category of content in your portfolio, list the primary traits or characteristics that you think are important. Next, construct a rating scale that describes the range of student performance that can occur for each trait. Figures 9.1 and 9.2 show how Keith did this for the essay writing content area.

Insert Figures 9.1 & 9.2

Figures 9.3 and 9.4 show examples from a math portfolio under the content category of problem solving. The teacher wants to measure the cognitive outcomes of knowledge base, cognitive strategies, communication, and reflection.

Insert Figures 9.3 & 9.4 about here

Once you design rubrics for each entry in the portfolio, you next design scoring criteria for the portfolio as a whole product. Some traits to consider
when developing a scoring mechanism for the entire portfolio are thoroughness, variety, growth or progress, overall quality, self-reflection, flexibility, organization, appearance, etc. Choose among these traits or include others and build five point rating scales for each characteristic.

Thus, the key to Step 5 is to do the following:

(1) For each cognitive skill and disposition in each content area, build your scoring rubrics.

(2) Put these on a form that allows you to include ratings of early drafts.

(3) Prepare a rating for the portfolio as a whole.

APPLICATION

In the space below prepare a rubric for one of the content areas identified in Step 4. Also, indicate the type of scale for rating the portfolio as a whole.

Rubric: __________________________________________

________________________________________________

-                                                                                      -

________________________________________________

-                                                                                      -

Scale: __________________________________________

________________________________________________

-                                                                                      -

9-20
Step 6: Developing a Procedure to Aggregate All Portfolio Ratings

For each content category that you include in the portfolio, learners will receive a score for each draft and the final product. You will have to decide how to aggregate these scores into a final score or grade for each content area and, then, the portfolio as a whole. Figures 9.2 and 9.4 are examples of a summative rating form in two content areas (essay and math) for one student. You will have one of these summative rating forms for each content area identified in Step 4. Thus, if you want a writing portfolio to include five areas of content (persuasive writing, dialogue, biography, criticism, commentary), you will have five summative rating forms, each of which rates drafts and final product.

As you can see in Figures 9.2 and 9.4, the teacher averaged the ratings for the two preliminary drafts and the final one. The next step is to develop a rule or procedure for combining these three scores into an overall score. One procedure would be to compute a simple average of the three scores. This method gives equal importance in the final score to the drafts and final product. Another procedure would be to assign greatest importance to the final product, lesser importance to the second draft and least importance to the first draft.

This is called weighting and we will discuss this process at greater length in the final chapter when we discuss developing a grading system. If and how you weight scores is up to you. You might seek input from learners and parents but there is no hard and fast rule about whether or which products in an area should be given more weight.

If you should decide to assign different importance or weight to the
products in a content area, then do the following:

(1) Decide on the weight in terms of a percentage, e.g., first draft counts 20%, second draft counts 30%, final draft counts 50% of final score. Make sure the percentages add up to 100%.

(2) Take the average score for each product and multiply that by the weight. In our example as shown in Figure 9.2, this would involve the following calculations:

Draft 1: \[3.50 \times 0.2 = 0.7\]
Draft 2: \[4.25 \times 0.3 = 1.3\]
Final: \[3.75 \times 0.5 = 1.9\]

(3) Add up these products (.7 + 1.3 + 1.9) and you get an overall score of 3.9 for the content area of essay writing. We will consider the meaning of this value shortly. (Had you not weighted, the average score would have been 3.8.)

Follow this same procedure for each content area. If you have five content areas in the portfolio, you will have five scores. Let's say that these are:

<table>
<thead>
<tr>
<th>Content Area</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essay</td>
<td>3.9</td>
</tr>
<tr>
<td>Dialogue</td>
<td>4.0</td>
</tr>
<tr>
<td>Criticism</td>
<td>2.5</td>
</tr>
<tr>
<td>Biography</td>
<td>3.8</td>
</tr>
<tr>
<td>Commentary</td>
<td>2.0</td>
</tr>
</tbody>
</table>

The next step is to decide how to aggregate these scores. Again, you can choose to weight, or not to weight. You may decide to involve learners and their parents in this decision. If you decide not to weight than the average
rating for all the content areas is 3.2 (rounded to the nearest decimal).

Finally, assign a rating to the portfolio as a whole. Let's say that the rating came out to be a 4.5. Now you must decide how to include this rating in the overall portfolio grade. If you take an unweighted average, you assign as much importance to that one rating as you did all the separate content ratings. That's probably not a good idea. Your average grade of 3.9 for the portfolio areas taken separately is a more reliable rating than your one rating of 4.5 for the whole portfolio. So, we recommend that you assign more weight to the former score than the latter--let's say 90% vs. 10%, which produces a final grade of:

\[
3.9 \times 0.9 = 3.51 \\
4.5 \times 1 = 4.45 (0.10) = 4.5
\]

Final Grade = 3.51 + 0.45 = 3.96 (versus 4.2 if you had you not weighted).

Now, let's consider what 3.96 (or 4.0 rounded) means in terms of the quality of the overall portfolio. Remember the distinction in Chapter 2 between measurement and evaluation? The value 4.0 is a measurement. It is a number that reflects the performance of the learner. But, how good is a 4.0 in indicating the competence of the learner? Making this decision involves evaluation.

Here is one way to assign meaning to our measurement of 4.0. Schools usually assign the following values to grades:

<table>
<thead>
<tr>
<th>Grading Schemes</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>90 - 100</td>
<td>A E</td>
</tr>
<tr>
<td>80 - 89</td>
<td>B S+</td>
</tr>
<tr>
<td>70 - 79</td>
<td>C S</td>
</tr>
<tr>
<td>60 - 69</td>
<td>D S-</td>
</tr>
<tr>
<td></td>
<td>Outstanding</td>
</tr>
<tr>
<td></td>
<td>Above average</td>
</tr>
<tr>
<td></td>
<td>Average</td>
</tr>
<tr>
<td></td>
<td>Below average</td>
</tr>
</tbody>
</table>

9-23
below 60 F N Failure, not at standard, etc.

When using five-point rating scales, we usually consider "3" as average, "1" as below standard, and "5" outstanding. Similarly, if you used a 7 point scale, a 3.5 would be average, ratings between 1 and 2 are below standard, and ratings between 6 and 7 are outstanding. So, one way to assign value to a 4.0 would be to link the traditional grading systems and their conventional meanings to scores on the rating scale. Pick a range of rating scale values that correspond to a letter or numerical grade in your school and link the two:

<table>
<thead>
<tr>
<th>Average Rating</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0-1.9</td>
<td>F, 50 - 59</td>
</tr>
<tr>
<td>2.0-2.5</td>
<td>D, 60-69</td>
</tr>
<tr>
<td>2.6-3.6</td>
<td>C, 70-79</td>
</tr>
<tr>
<td>3.6-4.3</td>
<td>B, 80-89</td>
</tr>
<tr>
<td>4.4-5.0</td>
<td>A, 90-100</td>
</tr>
</tbody>
</table>

If we use the above chart, a 4.0 would represent a grade of "B," a numerical grade somewhere between 80-89, or a grade of Satisfactory + (S+). Depending on factors that we will discuss in Chapter 11, you may want to add plus' and minus' to your grading system. Or, decide that a B gets a grade of 85, B- gets a grade of 80, and B+ a grade of 89. Making these decisions before you begin grading the portfolios and evaluating each portfolio using the same criteria helps minimize subjectivity.

In Chapter 11 we will discuss the role of the portfolio in determining the overall course grade.
APPLICATION

In the space below, describe a procedure to aggregate all portfolio ratings, and to assign grades to the completed portfolio. Decide how you would weight (1) drafts when computing a content area rating, (2) content area ratings when you average these, and (3) your rating of the whole portfolio with the average rating of the content areas.

1. Drafts: ______________________________________

____________________________________________________________________

2. Content areas: ______________________________________

____________________________________________________________________

3. Whole portfolio: ______________________________________

____________________________________________________________________

Now, construct a chart in the box below that shows how you will link the overall portfolio score with a grading scheme of your choosing.
ACT III, SCENE 1: FIFTH WEEK OF CLASS

During the past week Keith had his individual conferences with each student and finalized the grading scheme. His plan for this week was to tie up some loose ends.

Keith: Take out the forms we developed for grading each area of the portfolio. We have to decide when you turn in the work. Look at the class syllabus you got the first week of class and see the order in which you'll be learning the different writing styles. First, is persuasive writing which we already started. Then we cover editorials, and so on. So take the persuasive writing form, and let's write in the dates when the drafts are due and the final work.

Abebe: But, how do we know when we'll get them back so we can revise?

Keith: Good question! You'll get them back after three class days. OK? so let's get the first draft of your first sample of persuasive writing in by October 21st. Write this date on the form? Any more questions?
Step 7: Determining the Logistics

So far, you have accomplished these aspects of portfolio design:
(1) Specified the purpose of the portfolio
(2) Identified the cognitive skills it will reflect
(3) Decided who will help plan it.
(4) Decided what and how many products go in it.
(5) Specified the rubrics by which to score it.
(6) Developed a rating and grading scheme.

There are just a few details left.

Timelines. Your learners and their parents need to know exact dates when things are due. Just as Keith did above, link these dates with your syllabus. Point this out to your learners. This reinforces in your learner's minds the link between your teaching and what's required in the portfolio. Be prepared to revise some of your requirements. You may find that there's not enough time in the school year and not enough hours in a week for you to read all the drafts and products and get them back to your learners in a timely fashion.

How Products Are Turned In And Returned? Decide how you want your learners to turn in their products. At the start of class? Placed in a "In" basket? Secured in a folder or binder? Returned in an "Out" basket? How will late assignments be handled? How do absent learners submit and get back assignments? Will there be penalties for late assignments?

Where Final Products Are Kept? Decide where the final products
will be stored? Will it be the learners responsibility to keep them safely at home? Or, do you want to store them so that they can be assembled easily for a final parent conference and passed on to other teachers? Remember that the products may include video or audio tapes so a manila folder might not work. You may need boxes, filing cabinets, or closets.

**Who Has Access To The Portfolio?** Certainly you, learners, and parents have a right to see what's in it. But, do other students, current and future teachers, administrators? You might want learners (and their parents) to help make these decisions since it is their portfolio and their story.

Plan a Final Conference

Plan to have a final conference at the end of the year or term with individual learners and, if possible, their parents to discuss the portfolio and what it says about your learners' development and final achievement. Your learners can be responsible for conducting the conference, with a little preparation from you on how to do it. This final event can be a highly motivating force for your learners to produce an exemplary portfolio.

**APPLICATION**

In the space below describe how you will handle the logistical issues discussed above.

Timelines: _______________________________________________________

_______________________________________________________________

9-28
A PORTFOLIO DEVELOPMENT CHECKLIST

The following checklist will help you as you design and revise your portfolio assessment program.

1. What purpose(s) will your portfolio serve? (check any that apply)
   - Prepare a sample of best work for future teachers to see
   - Communicate to parents what's been learned

9-29
☐ Evaluate my teaching
☐ Assign course grades
☐ Create collections of favorite or best work
☐ Document achievement for alternative credit
☐ Submission to a college or employer
☐ To show growth in skill and dispositions
☐ Other ________________________________

2. What cognitive skills will be assessed by the individual entries?
   ☐ Cognitive strategies (specify) ________________________________
   ☐ Deep understanding (specify) ________________________________
   ☐ Communication (specify) ________________________________
   ☐ Metacognition (specify) ________________________________
   ☐ Procedural skills (specify) ________________________________
   ☐ Knowledge construction (specify) ________________________________
   ☐ Other ________________________________

3. What dispositions do you want the entries to reflect?
   ☐ Flexibility
   ☐ Persistence
   ☐ Acceptance of feedback
   ☐ Others (specify) ________________________________

4. What criteria or rubrics will you use to judge the extent to which these skills and dispositions were achieved?


5. In rating the portfolio as a whole, what things will you look for?
   ☐ Variety of entries
   ☐ Growth in reflection

9-30
6. What kind of scale will you construct to rate the overall portfolio?

7. How will you combine all your ratings into a final grade?

8. Who will be involved in the planning process?
   - Learners
   - Teacher
   - Parents

9. What content categories are included in the portfolio?

10. Will learners have a choice over content categories?
    - Yes
    - No

11. Who decides what samples to include in each content area?
    - Learner
    - Teacher
    - Parents

12. How many samples will be included in each area?
    - One
☐ Two
☐ More than two

13. Have you specified deadlines for the entries?
   ☐ Yes
   ☐ No

14. Have you developed forms to rate and summarize ratings for all drafts and final products?
   ☐ Yes (specify) __________________________________________
   ☐ No

15. What are your instructions for how work gets turned in and returned?
   ________________________________________________________
   ________________________________________________________

16. Where will the portfolios be kept and who has access to them?
   ☐ Where (specify) ________________________________________
   ☐ When (specify) _________________________________________

17. Who will plan, conduct, and attend the final conference?
   ☐ Learner
   ☐ Other teachers
   ☐ Parents
   ☐ Others (specify) _______________________________________
Summary

Portfolios are a means of communicating to parents, learners, and other teachers the level of authentic learning and performance that a learner has achieved. By utilizing actual tasks—projects, scripts, essays, research reports, demonstrations, models, etc.—the learner applies domain specific and domain general knowledge and understandings to exhibit the level of deep learning that has been acquired from your instruction. Portfolio assessment is often the best and sometimes the only method for gauging your learners' level of deep learning. But, planning and designing a portfolio assessment must be as systematic and methodical as constructing an objective test or essay exam. Each cognitive skill and disposition, for each content area must be identified and scoring rubrics development. This chapter has helped you accomplish this by showing you how to select the purpose of your portfolio, the cognitive skills and dispositions it will measure, the criteria by which you will judge degrees of proficiency and the scoring rubrics by which you will rate and quantify your judgments. In the next chapter we will turn to the social and collaborative skills your learners may need to successfully complete their portfolios and to succeed in as well as outside your classroom.

Activities
1. Identify three threats to the validity of a portfolio and indicate what you would do in the design of your portfolio to see that these threats are minimized.
2. Plan a real portfolio by answering each of the questions on the Portfolio Development Checklist. Check the boxes that apply and provide the necessary details, where requested,
3. Develop a portfolio rating form and summative rating form for the entries
in your portfolio using Figures 9.1 and 9.2 as a guide. Be sure to include definitions for all the scale alternatives (e.g. "1" to "5") being rated as illustrated in Figure 9.1.

4. Describe the procedure you will use to aggregate scores for all the portfolio ratings. By providing hypothetical ratings for the entries on your rating form, indicate with actual numbers and averages, how you will (1) calculate weights, (2) take the average score for each entry, (3) add up all the entries to get an overall score, and (4) assign a grade symbol (e.g. A-F) to the average score.

Suggested Reading


Gives an exceptionally thorough treatment of the dimensions of learning that portfolios can assess and presents numerous examples of scoring rubrics.
Essay Portfolio Rating Form

___ First Draft
___ Second Draft
___ Final Draft

To be Completed by Student:

1. Date submitted: _____

2. Briefly explain what this essay says about you.
   __________________________

   __________________________

3. What do you like best about this piece of writing? ________________

   __________________________

4. What do you want to improve on the next draft? ________________

   __________________________

5. If this is your final draft will you include this in your portfolio and
   why?

   __________________________

   __________________________

To be completed by teacher:

9-35
1. Quality of Reflection

<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>States very clearly what he/she likes most and least about the essay. Goes into much detail about how to improve the work.</td>
</tr>
<tr>
<td>4</td>
<td>States clearly what he/she likes and dislikes about the essay. Gives detail about how to improve the work.</td>
</tr>
<tr>
<td>3</td>
<td>States his/her likes and dislikes but could be clearer. Gives some detail about how the work will be improved.</td>
</tr>
<tr>
<td>2</td>
<td>Is vague about likes and dislikes. Gives few details about how essay will be improved.</td>
</tr>
<tr>
<td>1</td>
<td>No evidence of any reflection on the work.</td>
</tr>
</tbody>
</table>

2. Writing Conventions

<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>The use of writing conventions is very effective. No errors evident. These conventions are fluid and complex: spelling, punctuation, grammar usage, sentence structure.</td>
</tr>
<tr>
<td>4</td>
<td>The use of writing conventions is effective. Only minor errors evident. These conventions are nearly all effective: punctuation, grammar usage, sentence structure, spelling.</td>
</tr>
<tr>
<td>3</td>
<td>The use of writing conventions is somewhat effective. Errors don't interfere with meaning. These conventions are somewhat effective: punctuation grammar usage, sentence structure, spelling.</td>
</tr>
</tbody>
</table>
2 Errors in the use of writing conventions interfere with meaning. These conventions are limited and uneven: punctuation, grammar usage, sentence structure, spelling.

1 Major errors in the use of writing conventions obscures meaning. Lacks understanding of punctuation, grammar usage, sentence structure, spelling.

3. Organization

5 Clearly makes sense.

4 Makes sense.

3 Makes sense for the most part.

2 Attempted but does not make sense.

1 Does not make sense.

4. Planning (1st draft only)

5 Has clear idea of audience. Goals are very clear and explicit. An overall essay plan is evident.

4 Has idea of audience. Goals are clear and explicit. Has a plan for the essay.

3 Somewhat clear about the essay's audience. Goals are stated but somewhat vague. Plan for whole essay somewhat clear.

2 Vague about who the essay is for. Goals are unclear. No clear plan evident.

1 Writing shows no evidence of planning.
5. Quality of Revision (2nd draft only)

  5 Follows-up on all suggestions for revision.
     Revisions are a definite improvement.

  4 Follows-up on most suggestions for revision.
     Revisions improve on the previous draft.

  3 Addresses some but not all suggested revisions.
     Revisions are a slight improvement over earlier draft.

  2 Ignores most suggestions for revision. Revisions made do not improve the earlier draft.

  1 Made only a minimal attempt to revise if at all.

Sum of ratings:  

Average of ratings:  

Comments:  


Figure 9.1. Essay Portfolio Rating Form

9-38
Essay Summative Rating Form

(Attach to each completed essay)

___ Essay Sample One
___ Essay Sample Two

Student ______________________

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Draft 1</th>
<th>Draft 2</th>
<th>Final Draft</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rating</td>
<td>Rating</td>
<td>Rating</td>
</tr>
<tr>
<td>Reflection</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Conventions</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Organization</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Planning</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

Average 3.5 Average 4.25 Average 3.75

Teacher: Comments on final essay development __________________________

____________________________________________________________________

____________________________________________________________________

Student: Comments on final essay development __________________________

____________________________________________________________________

____________________________________________________________________

9-39
Parent: Comment on final essay development

Included in portfolio:  ____ Yes
                     ____ No

Figure 9.2. Essay Summative Rating Form
Math Problem Solving Portfolio Rating Form

Content Categories:

___ √ __ Problem solving
____ Numbers and operations
____ Estimation
____ Predictions

Sample ___ One
____ √ Two
____ Three

To be Completed by Student:

1. Date submitted: ______

2. What does this problem say about you as a problem solver?
   ____________________________
   ____________________________
   ____________________________

3. What do you like best about how you solved this problem?
   ____________________________
   ____________________________
   ____________________________

4. How will you improve your problem solving skill on the next problem?
   ____________________________
   ____________________________
   ____________________________

To be completed by teacher:

9-14
1. Quality of Reflection

<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Has excellent insight into his/her problem solving abilities and clear ideas of how to get better.</td>
</tr>
<tr>
<td>4</td>
<td>Has good insight into his/her problem solving abilities and some ideas of how to get better.</td>
</tr>
<tr>
<td>3</td>
<td>Reflects somewhat on problem solving strengths and needs. Has some idea of how to improve as a problem solver.</td>
</tr>
<tr>
<td>2</td>
<td>Seldom reflects on problem solving strengths and needs. Has little idea of how to improve as a problem solver.</td>
</tr>
<tr>
<td>1</td>
<td>Has no concept of him or her self as a problem solver.</td>
</tr>
</tbody>
</table>

2. Mathematical Knowledge

<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Shows deep understanding of the problems, math concepts and principles. Uses appropriate math terms and all calculations are correct.</td>
</tr>
<tr>
<td>4</td>
<td>Shows good understanding of math problems, concepts and principles. Uses appropriate math terms most of the time. Few computational errors.</td>
</tr>
<tr>
<td>3</td>
<td>Shows understanding of some of the problems, math concepts and principles. Uses some terms incorrectly. Contains some computation errors.</td>
</tr>
<tr>
<td>2</td>
<td>Errors in the use of many problems. Many terms used incorrectly.</td>
</tr>
</tbody>
</table>
1 Major errors in problems. Shows no understanding of math problems, concepts and principles.

3. Strategic Knowledge

5 Identifies all the important elements of the problem. Reflects an appropriate and systematic strategy for solving the problem; gives clear evidence of a solution process.

4 Identifies most of the important elements of the problem. Reflects an appropriate and systematic strategy for solving the problem and gives clear evidence of a solution process most of the time.

3 Identifies some important elements of the problem. Gives some evidence of a strategy to solve the problems but process is incomplete.

2 Identifies few important elements of the problem. Gives little evidence of a strategy to solve the problems and the process is unknown.

1 Uses irrelevant outside information. Copies parts of the problem; no attempt at solution.

4. Communication

5 Gives a complete response with a clear unambiguous explanation, includes diagrams and charts when they help clarify explanation; presents strong arguments that are logically developed.

4 Gives good response with fairly clear explanation, which includes some use of diagrams and charts; presents good arguments that are mostly but not always logically developed.

9-43
3 Explanations and descriptions of problem solution
   are somewhat clear but incomplete; makes some use
   of diagrams and examples to clarify points but
   arguments are incomplete.

2 Explanations and descriptions of problem solution
   are weak; makes little, if any, use of diagrams and
   examples to clarify points; arguments are seriously
   flawed.

1 Ineffective communication; diagrams misrepresent
   the problem; arguments have no sound premise.

Sum of ratings: __________

Average of ratings: __________

Comments: ______________________________________________________

__________________________________________

__________________________________________

Figure 9.3. Math Problem-solving Portfolio Rating Form
Math Problem Solving Summative Rating Form

(Attach to problem solving entries)

___ Essay Sample One
___ Essay Sample Two

Student _______________________

<table>
<thead>
<tr>
<th>Problem 1</th>
<th>Problem 2</th>
<th>Problem 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criteria</td>
<td>Rating</td>
<td>Criteria</td>
</tr>
<tr>
<td>Reflection</td>
<td>3</td>
<td>Reflection</td>
</tr>
<tr>
<td>Knowledge</td>
<td>2</td>
<td>Knowledge</td>
</tr>
<tr>
<td>Strategies</td>
<td>2</td>
<td>Strategies</td>
</tr>
<tr>
<td>Comm.</td>
<td>2</td>
<td>Comm.</td>
</tr>
</tbody>
</table>

Average 2.25 Average 2.75 Average 2.5

Teacher: Comments on problem solving ability and improvement:

_________________________________________________________________
_________________________________________________________________
_________________________________________________________________

Student: Comments on problem solving ability and improvement:

_________________________________________________________________
_________________________________________________________________
Parent: Comments on problem solving and improvement:

Figure 9.4. Math Problem Solving Summative Rating Form
Chapter 10
Assessing Collaborative Skills And Self-Determination

You have learned a lot about assessing what your learners know and can do. You can test for knowledge, performance skills, problem solving strategies, and deep understanding. Most of your assessment program will focus on these outcomes. But there are two other areas of learning that you should address if you want a thorough picture of your learners' accomplishments: collaborative skills and self-determination.

Cognitive scientists for the past two decades have urged educators to make classroom learning more of a joint cognitive venture among all classroom participants than a solitary enterprise (McGilly, 1994). They encourage teachers to use small group teaching formats to advance the learning of their students, in particular, when this learning involves problem solving and deep understanding. There is little question that such grouping arrangements lead to higher degrees of complex learning in comparison to whole group teaching methods (Cohen, 1994).

But, experts in cognitive learning also emphasize that the collaborative skills necessary for learning in small groups are not automatically acquired by some children. Skills like the ability to disagree, ask for help, initiate conversations, give directions, and to give and receive compliments, are essential for productive small group learning and must be taught. And, if you must teach them, then you also must assess their achievement. This chapter will help you do that.

Cognitive scientists also tell us that intrinsic motivation for learning is indispensable for knowledge construction and deep understanding. Intrinsic motivation is what impels learners to choose certain tasks, get energized
about them, and persist until they master them successfully, regardless of whether it brings an immediate reward. There are many perspectives on how such motivation develops (O'Neil & Drillings, 1994). We reviewed some of them in Chapter 4. In this chapter, we will focus on recent developments in an area called self-determination (Deci, et al., 1991).

Self-determination includes such capabilities as choice making, decision making, goal setting, self-observation, self-evaluation, and self-reinforcement. We have discussed some of these skills before when we used the term metacognition. Metacognition involves knowing about yourself as a learner. The difference between metacognitive skills and self-determination skills is that the latter implies not only a direction for learning but a passion, intensity, or drive for it that metacognition does not.

Specialists in intrinsic motivation tell us that learners aren't born with it. Rather, it is a cognitive orientation or disposition that develops out of experiences with learning new things. Classroom events and circumstances can kindle or stifle it. In this chapter we will review what promotes self-determination, discuss briefly how to teach it, and spend a good deal of time describing how to assess it.

COLLABORATIVE SKILLS

According to cognitive scientists, classroom instruction that focuses on knowledge construction, active learner involvement, and deep understanding has to occur in a social rather than solitary context. This is because group learning embodies a number of features that promote conceptual growth. First of all, it forces learners to adjust their thinking to that of others. When students have to think about the alternative viewpoints of group members, they have to elaborate and defend their own ideas and
debate the merits of their opinions to others. This promotes a deeper organization and understanding of their own knowledge.

Secondly, groups provide encouragement and support to their members for the skills that they display. Group instruction and group problem solving allow learners to assume different roles and responsibilities for learning, and acquire new skills. Group members promote the development of these skills in one another because they help the group accomplish its goals.

Finally, when students are allowed to work with others, they observe one another's thinking and working processes. As group members debate and discuss with one another, their thinking strategies become more explicit. Social learning research shows that children can learn good thinking strategies by observing how their peers reason and problem solve (Zimmerman, 1990).

Thus, any teacher who wants to help her learners become better thinkers will want to use small group learning formats some of the time. Cooperative learning is one such instructional format. There are a number of excellent texts that can serve as guides to using cooperative learning (Borich, 1996, Cohen, 1986, Johnson & Johnson, 1991). A principal feature of cooperative learning and other small group instruction formats is the explicit, structured teaching of collaborative skills.

Teaching Learners to Collaborate

The structured teaching of collaborative skills consists of four techniques: *modeling*, *role playing*, *performance feedback*, and *transfer of training* (McGinnis and Goldstein, 1984).
Modeling. Teachers that use modeling are applying a powerful technique for promoting learning. But, using it correctly involves more than demonstrating a skill and expecting that learners will imitate it. The expert use of modeling includes selecting models whose characteristics are appealing to those who are observing and learning. In other words, the more the observer identifies with the modeler, the more that is learned.

The modeler must be able to demonstrate the collaborative skills clearly. Whether these skills involve expressing opinions, giving feedback, or accepting criticism, the modeler must be able to demonstrate the behaviors in a clear and detailed fashion. This involves displaying the least difficult behaviors before the more difficult, giving sufficient practice to facilitate over learning, and using as few irrelevant gestures and expressions as possible.

Role playing. Students need to practice the collaborative skills that they observe others using. Modeling alone isn't enough to strengthen skilled performance. Teachers who use role playing skillfully give their learners choices over parts to play, praise or use other forms of recognition and reward for playing the part, elicit a strong commitment from the learner for the behavior he or she is role playing, and allow the student to improvise or be creative when acting out the collaborative skills.

Feedback. After playing a role, learners must know how well they did. Teachers can communicate about the quality of a learners performance by making suggestions for improvement, providing additional practice, coaching, and expressing praise and approval. Making videotapes of the role playing and reviewing the performance with the learner is an especially
effective way of giving feedback.

**Transfer.** Students who learn new collaborative skills through role playing often don't use these skills in actual small group learning contexts in the classroom. If you expect an automatic transfer of the learning you see during role playing to real classroom situations, you may be disappointed. You have to "teach for transfer." This usually involves making the role playing as realistic to classroom situations as possible, practicing with a variety of peers using a variety of classroom situations, and using materials, furniture, and other role playing props that are as like actual classroom conditions as possible.

A Taxonomy of Collaborative Skills

Learners need four clusters of collaborative skills to work well in small group learning formats: they must have (1) basic interaction skills; (2) getting along skills; (3) coaching skills; and (4) skills in fulfilling particular roles (Hoy and Gregg, 1994).

**Basic interaction skills.** Students in a group have to learn to like and respect one another. In order for this to happen, at a minimum, each group member needs the following skills:

1. Listening
2. Making eye-contact
3. Answering questions
4. Using the right voice
5. Making sense, and
6. Apologizing.

**Getting along skills.** Since the members of a group will be working together for a length of time, they must be able to sustain their respect and liking for one another. Some of the skills involved in getting along with one another are:

1. Taking turns
2. Sharing
3. Following rules
4. Assisting
5. Asking for help or a favor
6. Using polite words.

**Coaching skills.** When people work together to accomplish some goal, they have to give and receive both corrective feedback and encouragement. Some of the skills involved in doing this are:

1. Suggesting an action or activity
2. Giving and receiving compliments or praise
3. Being specific
4. Giving advice
5. Correcting and being corrected.

**Role fulfilling skills.** In productive small groups, each member has a role to play. The fulfilling of these roles creates positive interdependency and individual accountability. The former characteristic leads group
members to depend on one another for the achievement of group goals. The latter prevents any peer member from slacking off. Examples of group roles include:

1. summarizer
2. checker
3. researcher
4. runner
5. recorder
6. supporter
7. troubleshooter

Once you identify a necessary role in a small group, you define the responsibilities of the role and explain these to the respective group member.

For example, the responsibilities of the recorder might be to: write up the major product; require others to write conclusions; collect and organize the written contributions. A summarizer might paraphrase conclusions of the group, elicit agreement or disagreement from group members, present the groups position to the whole class.

ASSESSING COLLABORATIVE SKILLS

As with the assessment of cognitive learning skills, there are issues related to validity and reliability that you need to give serious thought to (See Chapter 4 for a review of these issues). Recall that validity concerns whether the behaviors that you choose to observe or rate as indicators of the collaborative skills are the most important or relevant ones. Reliability deals with whether you have enough samples of behavior to judge that the learner
possesses the skill, and whether your instrument for gathering data has clear procedures for doing so.

Having reflected on these two critical issues and taken steps to assure them, here is a guide to follow when developing your approach to assessing collaborative skills.

Step 1: Clarify Your Reasons for Assessing Collaborative Skills

There are number of reasons for assessing collaborative skills. You may, for example, want to rate social skills in order to incorporate this rating into a learners overall grade. Learners will be more motivated to learn and demonstrate collaborative skills if they are held accountable for performing them.

Or you may choose to assess these skills to track a learner's progress over the course of a semester or grade. You may not want to make the learning of collaborative skills a component of the course grade, but rather view it as a necessary skill that enhances how students experience school. In such a case, you might measure collaborative skills at several times during the year to give a learner and his parents feedback on skill development.

Finally, you may want to measure collaborative skills before you begin small group learning in order to help form your groups. A key characteristic of small group instruction is using what are called heterogeneous groups. This means that you want each group to represent a variety of ethnicities, social status, cognitive abilities, scholastic achievements, personalities, and social skills. Knowing something about your learners' collaborative skills ahead of time will allow you to set up groups where learners have varying degrees of ability to take responsibility, get along with others, coach, and give feedback.
APPLICATION

In the space below identify your reason(s) for wanting to assess collaborative skills.


Step 2: Describe the Assessment Context

In many ways, your choices about the assessment context for collaborative skills are similar to your choices for procedural skill contexts.

 Specify the setting and level of obtrusiveness. In Chapter 6 you had to decide whether to use natural or simulated settings to observe and rate procedural skills. Re-read that discussion. With the assessment of social skills, it would seem that natural settings are best. Do you agree?

Likewise, as with procedural skills, you must decide whether or not your learners will know that you are assessing their social skills. As you will recall, the important considerations to weigh when observing procedural skills related to anxiety and motivation is whether to inform your learners that you will be observing and rating their collaborative skills. Our preference would be to let group members know that you are going to be
observing and rating them. What's your preference and why?

**Specify who will observe and rate the collaborative skills.** When judging procedural skills, you have options of teacher ratings, self-ratings, peer ratings, other adult ratings, or any combination of these. There are reasons for using one in place of another. Review these considerations (see Chapter 6, pp. ) and decide who would be the most appropriate person to rate the collaborative skills of your learners. In your opinion, who would this be and why?

**Specify how much evidence is needed.** Here you will need to plan to collect enough evidence so that your conclusions as to your learners' collaborative skills are reliable. Review the considerations for weighing how much evidence is needed for rating procedural skills on pages ___ to ___ in Chapter 6.

Step 3: Identify and Describe Each Collaborative Skill

Decide which of the above collaborative skills you want to assess: basic interaction skills, getting along skills, coaching skills, role fulfilling skills. Or consult other guides for identifying and categorizing them. There are several curricula for teaching social skills such as "Skillstreaming the elementary school child: A guide for teaching prosocial skills," (McGinnis and Goldstein, 1984). There are also standardized, published tests of social skills (see Hoy and Gregg. 1994, Chapter 7) that suggest behaviors for you to focus on.

Once you have identified the collaborative skills, then it is important to describe exactly what they entail. If you feel that listening is an important
collaborative skill that you want to assess learners on, you need to describe what you mean by that term. For example, you may define "listening" for your particular learners as: Attends when someone else is speaking (looks at them, orients to them) and tries to understand what is said (i.e. may ask a question). Or "asking for help" might mean: The student realizes when he/she needs assistance and asks for it in a friendly manner.

These definitions will vary depending on the age of the learner, the nature of the small group instructional method, your own particular beliefs about what behaviors are and are not important as indicators of a collaborative skill, and so on. It is important in defining a social skill to have a clear picture of what you want your learners to do, or how you want them to behave when they are listening, giving a compliment, or suggesting an activity, and then writing this picture down in such a way that you can observe and rate it reliably.

Step 4: Design the Rating Plan

As you will recall from Chapter 6, you have four choices for the rating plan: checklists, rating scales, holistic scoring, anecdotal records. Review the strengths and limitations of each one and decide which plan to use for your learners. Here are examples of each system for rating collaborative skills:
COLLABORATIVE SKILLS CHECKLIST

Setting: ________________________ Date: _______

Directions: Place a check (√) in the appropriate box if you saw the student demonstrate this skill as the small group was working:

(Editor: Be sure art work is inserted on next two pages)

1. **Praising.** Student complimented the work or contribution of a peer.
2. **Listening.** Student attended to others when they were speaking and acknowledged that he/she heard what was said.
3. **Followed rules.** Student complied with all group rules.
4. **Assisting.** Student gave help when asked.

RATING SCALE FOR COLLABORATIVE SKILLS

Directions: Each of the statements below asks you to rate each group member on how well they did. In the boxes below each name, indicate how well they performed the skill using the following scoring code:

5 = Didn't miss an opportunity to do this
4 = Almost always did this

10-12
3 = Usually did this
2 = Seldom did this
1 = Never did this

1. Spoke to group members in an appropriate tone of voice.
2. Acknowledged compliments coming from peers.
3. Gave appropriate help when asked.
4. Told others when they did something good.
5. Clearly explained how to do something when asked.

HOLISTIC RATING OF COLLABORATIVE SKILLS

Directions: Rate the extent to which each group member demonstrated the following categories of skill using the following scale:

5 = Highly competent at these skills
4 = Pretty good in using these skills.
3 = Moderately good at these skills.
2 = Awkward in using these skills.
1 = Lacks these skills
Basic interaction skills. (The student shows that he likes and respects other members of the group.)

Getting along skills. (The student shows that she can make and keep friends in a group.)

Coaching. (The students shows that he can help and explain things to other group members)

Role fulfilling. (The student carries out his/her assigned responsibilities).

NARRATIVE SUMMARY OF GROUP COLLABORATIVE SKILLS

Group 1 2 3 4 5  

Date ____________________

I observed Group 3 for about 15 minutes today during the small reading group. They listened attentively to what others were saying but some did not volunteer information readily. But when asked to say or explain something, they did so very clearly and this is a significant improvement over a month ago. They still need to be more patient with one another especially when giving directions to peers who don't understand something. This group continues to be very responsible in carrying out their group assignment.

Step 5: Develop a Scoring Form for Recording your Ratings and Arrange the Collaborative Skills in an Appropriate Format.

We've given you some examples of recording formats above. Have a space on the form for recording name, place, date, and how long you
observed. It's a good idea to have a separate form for each small group rather than trying to get all class members' names on one form. We also recommend that you arrange the skills according to the cluster that they are best grouped in: for example, keep all getting along skills together, coaching skills, etc.

Step 6: Plan for Giving Feedback to Your Learners.

As we emphasized above when discussing teaching collaborative skills, feedback to learners on their performance is essential for learning and improvement. The more immediate you give this feedback, that is, the closer your feedback is to the time in which the skill was performed, the better the skill will be learned. Also consider videotaping or audio taping the interaction among group members. It is easier to point out problems in communication with others when you have something that learners can hear or see.

You must also decide whether to give this feedback individually or in a group. There are pros and cons to each. On the pro side, you are teaching group interaction skills. They are better learned in a group format and this includes feedback. On the con side, individual learners may be embarrassed in front of their peers when you point out certain limitations in their basic interaction and getting along skills.

SELF-DETERMINATION

In Chapter 2 we discussed motivation for classroom learning and some of the reasons for it's pronounced decline in many learners during the school year. We pointed out some of the things that teachers can do to enhance motivation among which was designing as assessment system that
is closely linked with instruction. We also described a shift in the past several decades from an emphasis on external forces as causes of motivation (i.e. the use of reinforcement and punishment to motivate learners) to a concern for internal or cognitive factors that rouse learners to higher degrees of effort.

As you will recall, "goal theory" is a recent advance in understanding academic motivation. It's main point is that learners who approach school tasks with the goal of mastery (as opposed to getting a good grade or being first in the class) have more positive attitudes towards school, seek out more challenging work, report using more effective problem solving strategies, and are more intrinsically motivated when compared to learners who lack such an orientation. An expression for describing these learners is that they are more "self-determined" (Deci, 1991).

Deci has devoted nearly a quarter-century to studying self-determination in school learners. He concludes that an attitude of self-determination to accomplish a particular goal grows out of three innate human needs: autonomy, relationship, and competence needs.

*Autonomy needs* refer to the ability to initiate and regulate one's own actions. Teacher's fulfill autonomy needs by giving learners choices over goals and activities to reach them. *Relationship needs* are innate requirements for secure and satisfying connections with peers, teachers, and parents. Classrooms that satisfy relationship needs promote high degrees of learner-learner interaction using such techniques as small-group learning that we have just talked about. Finally, *competence needs* involve the learner's knowledge of how to achieve certain goals and the skill for doing so. Teacher's meet learners' needs to feel competent by helping them set goals, make choices about goals, and monitor their achievement.
Self-determination, then, is a kind of intrinsic motivation characterized by a direction or focus on the goal of becoming academically competent, and an intense desire to get there by working in collaboration with peers and adults. Not all learners approach school learning with an attitude of self-determination. But, according to Deci, helping your learners become self-determined is one of the most significant gifts that you can give them now and in the future.

Defining Self-determination

The challenge in helping your learners develop an attitude of self-determination and assessing its growth is to be clear about what this behavior means. Or, in other words, what are the specific actions that can suggest that your learners possess this attribute?

Over the past decade there have been numerous attempts to define the behaviors that characterize learners who are self-determined (Agran, 1997; Wehmeyer, 1997; Deci, et al., 1991; Deci and Ryan, 1985). Here are some of the categories or classes of behaviors that have been identified: choice making, decision making, problem solving; goal setting; goal attainment; self-observation; self-monitoring; self-reinforcement; positive self-efficacy; self-awareness; self-knowledge.

In the interest of simplicity, we have organized these categories into four clusters:

1. Setting goals
2. Reaching goals
3. Observing performance
4. Rewarding oneself.
Setting goals. Goals are outcomes that the learner wants to achieve. Other words for goals are aspirations, purposes, aims, objectives, directions, and intentions. Examples of goals that learners may strive for are:

1. Learning how to solve geometric proofs.
2. Writing a poem for the school newspaper.
3. Completing a science project.
4. Constructing a topographical map.
5. Increasing reading speed and comprehension
6. Increasing sight word vocabulary

Learners who are self-determined, choose, select, or identify their own goals. Learners who are developing this capability need teachers to get them started.

Setting goals also involves identifying standards or criteria of performance. In the previous chapter, Keith not only asked his learners to identify goals for their writing samples, but also had them write down the qualities, criteria, or standards that these samples would reflect. Examples of criteria are the scoring rubrics that are depicted on pages ___ in Chapter 8 and on pages ___ in Chapter 9. For the examples of goals given above, criteria can be:

1. Increasing sight vocabulary by 50%.
2. Constructing a topographical map that includes mountains, ravines, streams, and rock outcroppings.
3. Increasing reading speed to 250 words per minute.
4. Learning how to solve geometric problems with 90% accuracy.
Self-determined learners should not only be able to identify goals for history, science, writing, or social behavior, but also identify degrees of quality for these goals.

Some of the behaviors or actions which suggest that your learners possess this aspect of self-determination are:

1. Sets goals that the learner genuinely wants to achieve.
2. Sets goals that focus on becoming skilled at a task.
3. Sets goals that are specific and measurable.
4. Sets goals that can be achieved in a specific time span.
5. Identifies times when goals will be achieved and sets intermediate times.
6. Identifies criteria of excellence for each goal.
7. Writes down goals and criteria.

Can you think of other actions which suggest that a learner possesses this aspect of self-determination?

**Reaching goals.** Reaching goals refers to the actions that learners take to achieve the goals that they identify for themselves. It includes such things as the amount of effort expended in reaching goals, adaptability in the face of obstacles, deep cognitive processing as opposed to surface or superficial attempts at understanding and learning, the use of strategies to accomplish tasks, seeking help as needed, and self-advocacy or assertiveness.

Some behaviors that suggest this aspect of self-determination are:

1. Persists at a task following unsuccessful performance.
2. Chooses alternate means to reach goals when one method fails.
3. Asks for help when needed.
4. Consults a variety of knowledge sources to achieve deep understanding.
5. Uses a clear strategy to reach the goal.
6. Works with those who can help him/her accomplish a task.
7. Is assertive with peers who unfairly interfere with goal attainment.
8. Helps keep others on task.

Can you add any others to this list? This and any of the above lists can be put in the form of a checklist

**Observing performance.** Students are self-determined observers when they keep a clear focus on goals and standards of performance and know when they are or are not making progress towards them. At first, they may have to rely on others to give them feedback on the quality of their work. But eventually they take more and more responsibility for noticing the discrepancy between where they are and where they want to be.

In some instances, observing performance requires learners to keep actual records of their work. In portfolio assessment, as you will recall, learners keep drafts of their work and monitor their progress. Often teachers ask students to graph or chart their spelling test scores, sight word vocabulary, foreign language vocabulary, completed assignments, punctuality, attendance, compliance with rules, etc. in order to track their progress in these areas.

Here are some behaviors which you can monitor in order to assess self-observation skills:
1. Keeps a record of progress towards goals.
2. Keeps daily (or weekly) graphs of performance.
3. Regularly checks records to keep track of progress.
4. Initiates discussions about records with teacher or coach.
5. Readily recalls specific aspects of performance when asked.
6. Comments on progress towards goal when asked.

Can you think of other learner actions that reflect self-observation skills?

**Rewarding oneself.** The final aspect of self-determination is self-reinforcement or the giving of rewards or recognition to oneself. When learners reach sought after goals or objectives, we want them to notice this and feel good about it. We want them to say to themselves things like, "Way to go!" "Nice work!" "I did it," etc. Certainly teachers, parents, and peers should recognize one another's efforts. But it's also important that learners depend on themselves as a source of encouragement. As we stated above, autonomy is a principle feature of feeling self-determined and self-reinforcement promotes independence.

Self-reinforcement can take a variety of forms. Learners can praise themselves by saying things covertly following achievement of goals. Or, they can treat themselves to something they wanted like an activity or something they wanted to buy. Whatever the reward is, learners should choose it and have say in the criteria to get it.

Behaviors which indicate that learners reward themselves are:
1. Expresses satisfaction when goals are achieved.
2. Can tell if a good job was done.
3. Only rewards himself/herself if goal is achieved.
4. Can postpone rewards until goal is accomplished.
5. Realistically evaluates performance.

What other behaviors can you observe which suggest that learners reward or self-reinforce themselves?

Assessing Self-determination

Learners aren't born with an attitude of intrinsic motivation or self-determination to do well in school. It is a direct result of experiences in school of the kind we talked about in Chapter 2. Thus, if you want your learners to have this trait, you must take specific steps to bring it about.

Very likely many of your learners will already be intrinsically motivated. But just as likely many will not. One of your goals for instruction may be to see your learners become more self-determined as the year progresses. So, knowing how to assess this attitude will help you meet this goal. As you will see, you already know about and have practiced the steps for doing so.
APPLICATIONS

Step 1: Be Clear about Your Purpose for Assessing Self-determination.

Just as with assessing procedural skills or collaborative skills, you must be clear at the outset why you want to assess self-determination. Most likely you want to do so because it is an attitude that affects lifelong learning and that all learners should possess. You want to assess it so that you can inform parents and learners themselves about this aspect of their school character. But, do you want to make self-determination a component of your grading system? If a learner shows no interest in mastering school tasks, will this be reflected in their grades? Or, do you simply want to note this and not consider it when filling out report cards?

We will discuss what should and should not be part of school grades in the next chapter. At this point, we simply want you to give thought to whether you want to systematically assess self-determination, and what you intend to do with the information once you get it.

In the space below explain why you want to assess self-determination and what you will do with the measurements once you get them.

____________________________________________________________________________

____________________________________________________________________________

____________________________________________________________________________

____________________________________________________________________________

____________________________________________________________________________

10-23
Step 2: Describe the Assessment Context.

The same considerations that apply to assessing social skills apply to assessing self-determination:

1. In what settings will you observe this attitude?
2. Will your learners be aware that you are observing them to make judgments about their self-determination?
3. What do you consider to be sufficient evidence of self-determination?

In the space below, answer the above questions.

1. __________________________________________________________
   __________________________________________________________
   __________________________________________________________

2. __________________________________________________________
   __________________________________________________________
   __________________________________________________________

3. __________________________________________________________
   __________________________________________________________
   __________________________________________________________

Step 3: Define What You Mean by Self-determination

We have discussed at some length Deci's perspective on self-determination and made some suggestions about its principal components and behavioral indicators. You may use this perspective on self-determination and/or add some of your own thinking about this attitude.
In the space below, identify the principal aspects of self-determination that you wish to assess and specify the behaviors that you will observe and measure as indicators of these.

Step 4: Design the Rating Plan

Choose from the following:

1. Checklists
2. Rating scales
3. Holistic scoring
4. Anecdotal records.

In the space below, choose one or more of the above systems for measuring self-determination and describe and indicate which one you will use.
Step 5: Develop Forms to Record Your Measurements of Self Determination

Examine examples of formats that we have provided in this chapter and other chapters. Arrange the indicators of self-determination according to some logical plan. In the box below, provide an example of the form(s) that you will use to record your ratings of self-determination.

Step 6: Plan for Giving Feedback to Your Learners and Their Parents

The nature of the feedback you give will depend largely on the purpose that you stated in Step 1. You will be talking to learners and their parents about some very personal characteristics. It is one thing to tell
learners that they need to acquire more skill at writing, problem solving, or reading comprehension. It is an entirely different to explain your perceptions of their motivation for school learning. The former information is universally regarded as part of the mission of classroom instruction. The latter isn't so widely included in a classroom assessment system. The sensitivity of this information should heighten your concerns about the validity and reliability of your measurements. In the space below, explain how you will give feedback to your learners and their parents about your judgments of self-determination.

Summary

Most of this book has been about measuring and evaluating the cognitive achievements of your learners. This chapter has focused on "affective" outcomes or achievements. Schools focus on these outcomes because they promote in important ways the learning of cognitive skills. Good thinking is enhanced when learners work in small groups. Collaborative skills are essential for the productive use of small group instruction. As the school year progresses, as children attend school year after year, they must begin to take greater responsibility for their own learning. An attitude of self-determination underlies this take-charge disposition.
Activities

1. Think of a collaborative skills activity for your grade or subject that you would like to teach. Using the taxonomies of collaborative skills provided, identify the basic interaction skills, getting along skills, coaching skills, and role-fulfilling skills you wish your learners to acquire.

2. Using the structured approach to the teaching of collaborative skills suggested in this chapter, indicate what you would do to: (1) model the collaborative skills identified in Question 1, (2) get learners to role play the skills you model, (3) provide learners feedback on their role playing, and (4) teach learners to transfer their collaborative skills to actual small group contexts?

3. Following the example provided in this chapter, create a checklist, a rating scale and a holistic scoring format for measuring the collaborative behaviors identified in Question 1. Divide up your intended behaviors across the three scoring formats, choosing the best format for the behavior being measured.

4. Identify at least five behaviors that measure self-determination. Place your behaviors in either a checklist, rating or holistic format, whichever you feel would best assess the behavior. Describe an anecdotal record that could be used to supplement your ratings of self-determination?

Suggested Reading

One of the best texts around for helping teachers identify social skills for their learners and ways to teach them.
Consider these comments made about grading by some first-year teachers:

Of all the paperwork, grading is the nitty-gritty of teaching for me. Students take their grade as the bottom line of your class. It is the end-all and be-all of the class. To me, a grade for a class is, or at least should be, a combination of ability, attitude, and effort. Put bluntly: How do you nail a kid who really tried with an F? Or how do you reward a lazy, snotty punk with an A? (Ryan, 1992, p. 4)

I have been amazed at the comments they make about grades. Those with A’s ask if they are flunking; others who rarely hand in assignments ask if they’ll get a B. They make no connection between their own efforts and the grade they receive...Of course, there is some truth in their view of the arbitrariness of grading. But I don't like the powerlessness it implies. "I don't give you your grade." I tell them. "You do!" (Ryan, 1992, p. 96)

Grading is still kind of a problem with me...I try not to play favorites (even though I have them)--I don't like S's personality...I do like J's, isn't it unfair? You want to be easier on someone you like. Or harder on someone you don't .... I wouldn't mind taking a class on grading. I think grading could be hit much harder in college...I just don't know what to do, kind of. (Bullough, 1989, p. 66)
Confusion, uncertainty, even fear can surround the practice of grading for beginning teachers. There are several reasons for this. First, there is little research supporting one grading practice over another (Thorndike et al., 1991). Teachers' choices of grading systems generally reflect their own values, past experiences with grades, and the norms and traditions of the school in which they teach. Second, grades become part of a learner's permanent record and have important consequences both at the time the grade is given and in the future. Teachers, therefore, are naturally cautious about how they assign grades and conscientious when they do so.

The goal of this chapter is to prepare you to develop and defend a grading plan. We will highlight significant decisions you will have to make in developing a grading plan and describe specific methods of calculating and assigning grades. Because there is little research that supports one grading approach over another, we won't describe a "best" method. But, we will provide some expert opinions and recommendations about grading practices and the reasons for them that will help you select the approach that makes sense to you and that you can explain and defend to learners and parents.

**Why Assign Grades?**

Why you assign grades to learners is one of the most important questions that you must answer. If you clearly articulate the purpose of your grades before deciding on how to assign them, many potential problems will be resolved. Before reading further, think about what your purpose for assigning a grade might be; then, write it in the space below:
Often the most important reason for assigning a grade is to communicate the extent to which learners achieve classroom goals. This reason is so obvious you may say that it's not worth mentioning. But, if it's obvious, why is it often ignored? When a teacher lowers a grade because a particular student shows "poor attitude," the grade no longer communicates just what the learner achieved. Which part of the grade represents achievement and which attitude? And, what about when a teacher decides to add points to a learner's grade because she or he "tried hard," or, when a teacher raises a grade because the student made a lot of progress. Which part of the grade represents achievement and which progress or effort? How would next year's teacher interpret a D grade assigned to reflect some achievement and a lot of progress compared to another learner's F grade assigned to reflect only achievement?

The principle that grades should be assigned to communicate what a student learned, while obvious, is often ignored in practice. Letter and number grades may communicate many things: achievement, effort, attitude, progress, the teacher's personal feelings about the learner, and more. If you want your grades to tell learners, parents and other teachers how well a learner has mastered your objectives, then your grades should communicate your students' achievement.
Cultivating a Grading Philosophy

You will face many decisions on the way to developing a grading plan. In this section we will highlight some of the questions you must consider before assigning grades, and, afterward, provide you with the tools to answer them.

1. What meaning should each report card grade convey? The choices are many: Should Marva's grade of 85 in eighth grade science tell how much information she has acquired, how much she has learned relative to her classmates, how much effort she put into the class, her class attendance and punctuality, how much progress she made from the start of the term, how much she contributed to class discussions, her appreciation for science, or how well she got along with her peers and the teacher? An "85" can't tell all of these things. What will it tell learners, parents, future teachers, or college admission counselors?

2. How should class grades be distributed? It is rare to find a school district with a written policy on how many students can receive any given grade. So you must make your own decision about what percentage of your learners get A's, B's, C's, etc. Is it OK for 90% of your learners to get A's? Should some students fail? Should your average grade be a B or a C? Were you to graph the distribution of grades, should it resemble a "bell-curve?"

3. What components go into a final grade? Students are asked to do many things in a classroom: complete class work, homework, projects, turn in rough drafts, keep a notebook or diary, take tests and pop quizzes and make oral presentations. Do all these go into a final grade? Is there a
minimum or maximum number of grade components?

4. How do you establish the importance of each component in the final grade? Should all of the components identified above carry equal weight in your final grade? Or are some more important than others? Should the five pop quizzes count as much as a chapter test? What percentage of the grade should depend on homework? How do you decide which should be more important and how should this be reflected in the final grade?

5. How should you assign the grades? Once you have combined the grade components into a final score based on the decisions above, you must specify a final grade for each learner? How you do this depends on the conclusions you draw about the meaning of a grade. If you decide that grades should tell what a student has learned in an absolute and not relative (comparison with other learners) sense, then you should not be concerned with how many learners get what grade. But there are several methods of assigning absolute grades? Which should you use?

6. Should students be given extra opportunities to raise their grades? What teachers haven’t had students come to them at the end of the term and ask for some special consideration regarding their grade. Sometimes just a few points affects whether a student goes to summer school, receives an award, passes the course, or earns a scholarship. What should you do in such cases? Give an extra assignment, have them take a "make-up" test, be a nice person and just give the extra points?
APPLICATION

Each of these issues will be discussed as we present the steps to developing a grading plan. Before going further, read each question again and make some decisions about what you would do?

1. What will be my purpose for giving a report card grade?

2. How will I distribute grades across the class?

3. What components of performance will I consider for the final grade?

4. How will I weight the importance of each component?

5. Will my grades be absolute or relative measures of achievement?
6. Will I give extra opportunities to raise a grade?

CONSTRUCTING A GRADING PLAN

In this section we present the important stages to follow when constructing a grading system. Even though we have numbered them sequentially, anticipate that you will have to revise an earlier decision after resolving a later one. Developing a grading plan is more like making Texas chili than assembling the parts of something you bought at Wal-Mart. Your final outcome will reflect some facts but also a lot of judgment, insight and experience.

Step 1: Check School District Policy

Every school district and most schools will have a policy regarding grades. The policies will likely indicate the grade symbols that you can use (e.g., A-F; 0-100), whether a final exam is required, what percentage of the final grade can be based on the final exam, etc. There may be detailed rules but more likely there will be general statements that are closer to a philosophy than a rubric, as shown in Figure 11.1.

Insert Figure 11.1 about here

You are bound to follow the district grading system. Where conflicts
arise between your approach and the district's or school's, discuss this with your principal or department head. Even if there is no obvious inconsistency between how you choose to assign grades and district guidelines, it is always helpful to discuss your grading approach with other teachers. This will help identify the norms for grading that exist in your school and prevent some potentially awkward situations later.

Step 2: Decide What You Want Each Grade Symbol to Mean

First, let's review the two basic classifications of symbols that are traditionally used in most schools: letter symbols and number symbols.

**Letter symbols.** A-F grades are the most widely used of the letter symbol systems. When '+' and '-' designations are added to these letters, the A-F system allows the teacher to judge 13 degrees of learning: A+, A, A-, B+, B, B-, C+, C, C-, D+, D, D-, F. We will return to the difficulty of discriminating 13 levels of learning later.

Many elementary schools use E (excellent), S+ (above expectations), S (satisfactory), S- (below expectations), N (needs improvement). This is a five degree or five level grading system. Often it is combined with another symbol that indicates whether the learner is working below (1), at (2), or above (3) grade level. Thus, a grade of S-3 indicates that the child is doing above grade level work at a satisfactory level.

Teachers who judge that the special nature of their class objectives require a pass-fail grading system use P-F or M-I (Mastery, Incomplete) letter systems.

**Numerical symbols.** The two numerical systems are the 0-100
approach, and the grade point average. With the former the teacher assigns scores between 0-100 to all components of the grading system: tests, projects, quizzes, notebooks, homework and classwork assignments, etc. These scores are averaged at the end of the grading period and this average is the grade reported on the report card. Although the teacher can record a grade anywhere in the 0-100 range, some school grading programs treat all grades below 60 as a 60. So, in effect, the 0-100 system is a 60-100 system.

Grade point average (GPA) systems assign a numerical rating to the letter grades that are reported by the teacher. Thus, an A = 4, B = 3, C = 2, D = 1, F = 0. These numbers are averaged for all subjects and the student receives a grade point average.

**Limitations of the symbol systems.** The principal limitation of any grading system that requires the teacher to assign one number or one letter to represent course learning is that one symbol can only convey one meaning. But how can we do justice to all that a student learns in a classroom with one symbol? In the typical class students acquire skills, knowledge, and dispositions about learning. One symbol cannot do justice to the different degrees of learning a student acquires across all learning outcomes. This limitation is an inherent feature of any symbol system. Below we will suggest ways of dealing with it.

Another limitation of the different symbol systems is that they require teachers to make either too few or too many judgments about levels of learning. For example, the 13 level A-F system assumes that a teacher can detect 13 degrees of performance. 0-100 or 60-100 point systems put nearly impossible demands on a teacher's ability to distinguish different levels of learning. At the other
extreme, Pass-Fail or Mastery-Incomplete systems may fail to do justice to the varying degrees of learning that occur in a classroom. Most measurement specialists recommend that a grading system include no fewer than 5 and no more than 10 degrees of quality.

 Assigning meaning to the symbols. Grade symbols typically describe one or more of the following:

(1) Achievement in comparison to a relative or absolute standard.
(2) How much effort a learner put into the class.
(3) How much growth in achievement occurred over the term or semester.

Grades that make relative comparisons are based on a student’s position in the class compared to the performance of other learners. For example, a grade of C is given to learners who achieve scores similar to most students on tests, papers, and other aspects of the class grading system. Learners whose class performance is well above the performance of most other learners receive A grades.

Grades that make absolute comparisons are based on the degree to which a learner meets a certain predetermined standard of performance. For example, a grade of C would be given to learners who achieve scores between 70 and 80 on all tests and ratings of 3 (on a 1-5 rating scale) on all papers and projects. Learners who earn 90 or better on all tests and scores of 5 on papers and projects receive a grade of A. Figure 11.2 describes two grading systems that assign relative and absolute meanings to A-F symbols.
Whether you choose an absolute or relative standard for your grades will affect the type of tests you construct or select as well as how you teach. Choosing grade symbols that reflect absolute standards require that you use measurement tools that are performance, mastery, or criterion-referenced (i.e., the type of measures that you learned about in this text). Giving grades meaning relative to how others performed requires that you use norm-referenced measurement tools and scoring systems that distribute grades in the form of a "bell-shaped" curve.

Teachers often choose to have the grade describe not only what the learner achieved but also the extent of the learner's effort. They do this by assigning a certain number of points (if using a 0-100 grading system) or parts of letter grades (raising a B grade to a B+ or lowering an A grade to an A-) based on their perception of how hard the student tried. Ideally, ratings of effort and ratings of achievement should be separated on the report card. Otherwise, a parent or future teacher will be unable to accurately interpret what the single grade means.

Finally, some grading systems incorporate growth or progress in addition to achievement and effort into the overall meaning of the grade. A distinction can be made between how much a student learned over time and level of learning attained at the end of the class. But, there are problems associated with efforts to base a grade's meaning on growth. An obvious one is that learners who enter the class with a high degree of skill have less room
to improve than those who enter with low skill levels.

Generally, grades should reflect the student's level of achievement at the end of the course and not the progress made across the term. While parents and their children are concerned about progress and growth, of more importance to them is the level of learning represented by the grade and whether this level is sufficient to allow the student to succeed in the next learning experience.

In summary, you should plan to have one meaning (e.g., achievement of course goals and objectives) for one grade. We also recommend that once you determine the meaning that you wish to assign to a grade, that you reflect on whether the meaning is consistent with your approach to teaching and assessment. For example, an obvious inconsistency would be if a teacher assigns grades with the intention of communicating absolute— or mastery—achievement, but uses norm-referenced tests, does not allow learners multiple opportunities to improve performance, and emphasizes competition, comparison, and ability over subject mastery.

Step 3: Distinguish Reporting and Grading Factors

You may feel uneasy about the "one symbol-one meaning" rule proposed in Step 2. Although incorporating behaviors like growth, effort, attitude, conduct, and attendance into a grade hampers the interpretation of achievement, these factors, nevertheless, are important aspects of a learner's overall classroom behavior. So, how do teachers communicate their judgments of these other behavioral components of a learner's classroom performance?

Teachers acquire a lot of information about the learning habits of their students during the course of a school year or semester: interests, conduct,
attendance, motivation, study skills, social skills, effort, persistence, etc. Parents, learners, school psychologists, and future teachers are interested in their observations of these behaviors. Thus, all of these variables make up a set of information called reporting variables. But, only some of these reporting variables are consistent with the meaning that a teacher assigns to his or her grades at the end of the grading period, called grading variables. One method of communicating these other behaviorial components of learning is to distinguish at the start of the school term the qualitative information (e.g., narrative statements) you want to report to learners, parents, and other teachers from the quantitative information (e.g. test, homework and portfolio scores, etc.) you want to use in composing a final grade.

As a teacher, you have a variety of ways to communicate reporting variables: comments on report cards, narrative statements, conferences with parents, letters of reference, etc. But, when it comes to grading variables you will have to use the grading symbols required by your school district. And, the grading symbols you assign to learners should be consistent with the achievement of your course objectives. Thus, as you work out your grading plan, identify the larger set of variables you want to report to learners, parents, and future teachers, and the subset of these variables you will use as a basis for your grades. Some grading variables that appear to be relevant for assigning grades often are not. For example, how do you interpret a biology grade where 10% depended on writing mechanics? In such a case, a student’s mastery of biology objectives may deserve a grade of B but poor spelling, organization, and writing mechanics dropped it to a C. Should we lower grades because of late papers or give learners an F for cheating on tests? Some teachers give learners a grade of F if they miss a test because of
an unexcused absence. In some cases such actions can distort the meaning of a grade intended to communicate mastery of course objectives.

How do you hold students accountable for attendance, effort, preparation for class, punctuality, etc. if the grading system doesn't allow for these. Behaviors related to attendance, cheating, punctuality, effort can often be addressed by the school's disciplinary policy. There are usually consequences for academic dishonesty, failure to come to class on time, or for skipping class. Check with other teachers and your school's disciplinary policy about how to hold learners accountable for these actions.

With respect to effort, a suggestion comes from an experienced high school teacher. On the first day of class she goes over her grading plan and tells her learners that just because they earn an A doesn't mean that she would write a letter of recommendation for them for a job or college. She makes it clear that she will only write letters of reference for those learners who demonstrate both excellent learning and good conduct and effort. Her learners get the message that all their actions in class have consequences. Some of these consequences are delivered in the form of grades. Others aren't. A similar procedure would be to identify a list of privileges for high effort students, as a means of communicating and rewarding the importance of these behaviors.

Step 4: Identify the Components of Your Grade

When you complete the previous two steps, you will have made two important decisions. Let's assume that you chose to make your grade symbols indicate achievement of your learning objectives and that you have identified the information you want to either qualitatively report to students and parents or use in determining a final numerical grade.
Your set of reporting variables can include practice tests, rough drafts of lab reports, daily homework assignments, in-class work, written answers to oral questions. Some of these variables are best viewed as activities that students complete as practice exercises. They may be reported but not included in a final grade. The final grade indicating mastery of course objectives represents those components that indicate achievement at the end of a unit of instruction: final drafts of portfolios, performance tests, knowledge tests, projects, completed and polished lab reports, etc.

How many examples (e.g., assignments, reports, tests) of each component to include in your grade will depend on the length of the grading period. One test in a six- or nine-week grading period is probably too little on which to base a grade. In general, the more information you have, the more likely your grades will represent your learners' true achievement levels.

This is the time to consider how you will treat group projects and cooperative learning activities as a component of your grading system. Will you give all group members the same score or try to distinguish individual contributions?

For example, you can have individuals rate each other on a five-point scale measuring the active involvement of each teammate in the group process, the average of which could be a score for individual effort. Or, you can rate the group's end product (or groups could rate one another and the average taken) on a five-point scale, providing two scores. Sample scales for measuring group and individual effort are illustrated in Figure 11.3.

Insert Figure 11.3 about here
Scores for these scales could be recorded independently of one another (e.g., individual effort = 4, group product = 5) or as a ratio (e.g., 4/5 = .80). If the ratio method is chosen each individual in the group would be given the same group score, either determined by you or by averaging each group's evaluation of one another. Ratios smaller than 1.0 indicate that the group product exceeded this individual's contribution to the group.

You may also want to try other types of group grades, such as:

1. Averaging of individual scores to determine the group grade.
2. Assigning all group members the average of the highest (or lowest) half of the members' scores.
3. Averaging an individual's score with the group score (for example, averaging an individual score of 4 with a group score of 5: \(4 + 5 = 9/2 = 4.5\)).
4. Adding points to the group score for each active participant within the group (or subtracting points from the group score for each nonparticipant), to be determined by the teacher.

Step 5: Decide How Much Weight Each Component Will Have in the Final Grade

Figure 11.4 shows the components of an eighth grade Life Science class and the weight or importance of the component expressed as a percentage of the final grade.

Insert Figure 11.4 about here
This teacher has decided that certain aspects of her grading plan will be more important for determining the final grade than others. How did she go about making this decision? There are three considerations: (1) the importance of the component as indicated by the number or extent of class goals and objectives that it measures; (2) the uniqueness of the objectives or goals measured by different components; and (3) the reliability of the scores.

The most important consideration in determining the weight of a component is the number of learning goals and objectives it covers. For example, let's say that when you plan your eighth grade science class you have a total of 12 learning objectives that you are working towards: three refer to developing the knowledge base, four relate to deep understanding, two deal with clear communication, and three with problem solving strategies. Thus, measures of deep understanding would carry most weight and clear communication least.

With respect to uniqueness, you may use both weekly homework assignments and objective tests to assess the knowledge base. In such a case, these components together should carry less weight than a performance test and portfolio designed to measure deep understanding.

Finally, some of your assessment techniques can be scored more reliably than others. Objective tests, for example, are usually more reliably scored than ratings of social skills. As a rule you should give more weight to the most reliably scored components of your grading system. However, this should not be at the expense of validity. In other words, if your objectives for a class focus more on deep understanding than acquisition of knowledge and procedural skills, weight the former component more than the latter.

Step 6: Determine How the Components Will Be Combined into a Final
Grade

How many points should a test, quiz, homework assignment, portfolio, performance test, class worksheet, etc., be worth when combining all component scores into a single score or grade? One approach is to score everything on a 100 point scale, average the scores for each component (e.g., average the scores of three objective tests), multiply each component average by the weight assigned, add up the weighted component points to arrive at the final grade. Figure 11.5 shows this process.

Insert Figure 11.5 about here

This is the most commonly used procedure for combining the various sub-component and component scores into a final grade. The principal problem with this procedure, however, is that it requires you to score every grading variable on a 0 - 100 point scale. For objective tests that have 50 items where you assign 2 points to each item, or 4 points per item on a 25 item test, this doesn't present a problem. But, consider a five question pop quiz. Grading this on the basis of 100 points by assigning 20 points to each question would presume you could distinguish a score of, say, 19 from a score of 20, or a score of 8 from a score of 9.

What about a homework assignment that asks the student to answer three essay questions at the end of a chapter? On what basis do you assign 100 points to this task, or to a two question essay test? How do you reliably distinguish 50 degrees of quality in an answer? Grading a book report on the basis of 100 points assumes that you can reliably distinguish an 85 from an 86, a 94 from a 95. Everything we know about a scorer's ability to detect
degrees of quality in a performance suggests that, at best, only 7 to 10 degrees of quality can be discriminated reliably.

Thus, the question "How many points should a grading variable be worth?" is best answered by, "As many points as can be reliably distinguished." 100 point grade combination systems may tax the ability of any teacher to reliably discriminate 100 degrees of communication skill, problem solving skill, depth of understanding, or knowledge.

The solution is to use a "percentage of total point system." Percentage of total point systems involve four steps:

1. Decide the components of your grading plan and assign each component a weight. A weight is the percentage of total points a particular component carries. Figure 11.4 displayed components and weights for a grading plan for an 8th grade Life Science class.

2. Record the actual points each student earned out of the number of points possible. Leave a column for totals. (See the sample grade record in Figure 11.6.) Each component and each separate assignment has varying numbers of possible points that can be earned. Assign points to each component based on the complexity of the required performance, the length of the assignment, and your perception of your ability to assign reliable ratings.

   Insert Figure 11.6 about here

3. Total the actual points earned for each component and divide by the maximum points possible. The results represent the percentage of points earned for each particular component. Thus, in our example
from Figure 11.6, Marvin and Darnell earned the following points and totals:

<table>
<thead>
<tr>
<th></th>
<th>Marvin</th>
<th>Darnell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework</td>
<td>50/70 = 71%</td>
<td>55/70 = 79%</td>
</tr>
<tr>
<td>Objective tests</td>
<td>45/60 = 75%</td>
<td>35/60 = 58%</td>
</tr>
<tr>
<td>Performance tests</td>
<td>33/40 = 83%</td>
<td>39/40 = 98%</td>
</tr>
<tr>
<td>Portfolio</td>
<td>18/20 = 90%</td>
<td>15/20 = 75%</td>
</tr>
<tr>
<td>Classwork</td>
<td>39/50 = 78%</td>
<td>37/50 = 74%</td>
</tr>
<tr>
<td>Notebook</td>
<td>5/10 = 50%</td>
<td>8/10 = 80%</td>
</tr>
</tbody>
</table>

4. Multiply each of these percentages by the weights assigned, as shown below:

<table>
<thead>
<tr>
<th></th>
<th>Marvin</th>
<th>Darnell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework</td>
<td>71 x .15 = 10.6</td>
<td>79 x .15 = 11.8</td>
</tr>
<tr>
<td>Objective tests</td>
<td>75 x .20 = 15</td>
<td>58 x .20 = 11.6</td>
</tr>
<tr>
<td>Performance tests</td>
<td>83 x .25 = 20.7</td>
<td>98 x .25 = 24.5</td>
</tr>
<tr>
<td>Portfolio</td>
<td>90 x .20 = 18</td>
<td>75 x .20 = 15</td>
</tr>
<tr>
<td>Classwork</td>
<td>78 x .10 = 7.8</td>
<td>74 x .10 = 7.4</td>
</tr>
<tr>
<td>Notebook</td>
<td>50 x .10 = 5</td>
<td>80 x .10 = 8</td>
</tr>
<tr>
<td>Totals</td>
<td>77.1</td>
<td>78.3</td>
</tr>
</tbody>
</table>

5. Record the grade either as a letter grade (A = 90-100 percent, etc.) or as the percentage itself, depending on your school's grading system.

Step 7: Choose a Method for Assigning Grades
Once you have computed the total number of points that a learner has accumulated across the various grading variables or components, you then must assign the grade that appears on the report card. If the report card is one where number grades are recorded, then the grade that a learner receives in a subject area is the number you recorded in Step 6. But, more likely you will have to associate this number with some type of letter grade using an A-F or similar (e.g., Excellent, Satisfactory, Needs Improvement) symbol system.

There are several relative and absolute grading methods of assigning letter grades. The former approaches include the grading-on-the-curve-method, the gap-method and the standard deviation method (Frisbie & Waltman, 1992). The procedures for using these relative methods are somewhat technical and beyond the scope of this book. These procedures also can sometimes be inconsistent with the philosophy of motivation, teaching, and learning necessary for developing an authentic assessment system. The absolute grading methods are more consistent with this text's approach to learning assessment. We will describe two of them.

The fixed percentage method of assigning grades. Teachers who use this method establish fixed ranges of cumulative grade scores developed in Step 6 as the basis for assigning the final letter grades. For example, a common set of ranges is: $A^+ = 95-100$; $A = 90-94$; $B^+ = 85-89$; $B = 80-84$, etc. These ranges are established at the start of the term. Thus, a learner whose cumulative point score computed in Step 6 was an 84 would receive a grade of B on his or her report card.

This is called an "absolute" method of assigning grades because the results of Step 6 determine the letter grade and nothing else. There are no limits set on how many A's, B's, or F's the class can receive. Theoretically,
every learner could receive an A or F, although this would be unlikely.

The principal drawback to this method of assigning grades is that the teacher does not know ahead of time how difficult or easy his or her assessment tasks are. Consequently, in the case of difficult tests, no learner would receive an A grade. When a teacher notices that all or almost all of her learners are earning no higher than B grades, she may adjust her ranges or add points to everyone's scores to get more A grades. This process diminishes the interpretability of the grade.

Another problem with fixed-percentage approaches is the arbitrary nature of the cut-off scores. There is no logical reason why an 89 is a B+ and a 90 is an A. Although only 1 point separates the A and B+ student, these two grades have vastly different culturally derived meanings for learners, parents, and future teachers. Thus, the grades themselves and the ranges of points that they encompass may have little meaning in terms of mastery of course content.

**The total point method.** Some teachers, rather than decide ahead of time the score ranges for a letter grade, accumulate points earned by their learners on all grade components and assign grades to the point total at the end of the grading period. Advance planning is needed for this method of assigning grades. First, the teacher must decide which components go into the final grade, how many scores for each component there will be, and the maximum number of points that each component and sub-component represents. This is decided before tests, papers, projects, or portfolios are developed and before the scoring criteria for the performance assessments have been established.

For example, a teacher may decide to use three tests of 20 points each
(60 points), five quizzes of 5 points each (25 points), one project worth 25 points, a portfolio worth 40 points, a notebook worth 10 points, and six weekly assignments worth 10 points each (60 points). Thus, the total number of points that can be earned is 220 points.

Next the teacher decides that she will assign A grades to learners who earn 90% of the total points, B grades to those earning 80% of points, etc. This produces score ranges of A = 198-220; B = 176-197; C = 154-175; D = 132-174; F = 110-131. The cut-off points are as arbitrary and may be as meaningless in terms of actual subject mastery as with the fixed-percentage method.

This system also has the limitation of the unknown difficulty levels of tests as does the fixed-percentage method. Some teachers avoid this limitation by using the average of the point totals earned by the top three students as the grade reference point. For example, if the top three learners earned total point scores of 200, 210, and 190 the teacher would average these (200) and use this total to set the ranges. Thus, 90% of 200 would set up a 180-220 range for A, 80% of 200 establishes an 160-179 range for B, etc. This procedure controls somewhat for difficulty level, but makes the interpretation of an A grade questionable in terms of absolute learning criteria.

Another drawback to this system is that the teacher must decide ahead of time how many points all tests are worth. For example, he decides that a unit test would be worth 20 points. As he starts building the test he wants to ask 12 objective type questions worth 2 points each and 3 short essays worth 5 points each. But, this exceeds the allowable point total by 9 points. So he takes out two objective-type questions and one essay to get the correct point total. But, in doing this he also deletes questions that are important to assess
specific content objectives.

Regardless of which method of assigning grades you select, or which method your school system will require that you adopt, you will not eliminate problems of subjectivity and interpretability, particularly when the report card requires that one symbol be assigned to summarize course achievement. The important point is to recognize the limitations of whatever grading system you develop.

Step 7: Decide What to Do about Borderline Scores

Expect to have learners whose point totals are one or two points away from the next higher grade category. Thus, a few points will make a difference between passing and failing a course, getting an A or B grade, getting a particular award, etc. There is a tendency to give learners extra credit work to make up the few points that they need, or even to give them the extra points to raise the grade.

Here are some questions to consider when deciding how to deal with borderline cases:

1. How many points below a grade cutoff must someone be to be considered borderline?
2. Will every borderline learner be given the opportunity to raise their score or just those assertive enough to ask you to do so?
3. If you give extra-credit work, how closely will this be related to the objectives or goals that the learner failed to master on your tests, projects, or portfolios?
4. Will the motivation of the learner be a factor in your decision to reconsider a borderline grade? In other words, will every
borderline learner have the opportunity to raise their grade, or just those whom you judge to be deserving of it?

Below is a checklist of questions to ask after you've developed your grading plan:
GRADING PLAN CHECKLIST

Check any that apply.

1. Have I checked my school district's, school's, or department's policy on grading?
   - If yes, specify where ________________________________

2. What symbol system does my school or school district use?
   - A-F
   - E, S+, S, S-, N (needs improvement)
   - M (mastery), I (incomplete)
   - 0-100
   - Other (specify) ________________________________

3. What behaviors will I assign a grade?
   - Achievement
   - Growth
   - Effort
   - Other (specify) ________________________________

4. How will I separate effort and growth from achievement in my grading plan?
   _____________________________________________
   _____________________________________________
   _____________________________________________
   _____________________________________________

5. What type of comparisons do I want my grades to represent?
   - Criterion-referenced (absolute comparisons)
6. Is the type of comparison I have chosen consistent with my approach to teaching? (explain why)

7. What behaviors have I decided to report but not grade?
   - Growth
   - Effort/persistence
   - Preparation
   - Conduct
   - Attendance
   - Group effort
   - Others (specify)

8. For those behaviors that I will not grade, how will I report and reward them? (specify)

9. What will be the components of my grading plan, their weights, and number of scores for each:

<table>
<thead>
<tr>
<th>Components</th>
<th>Weight</th>
<th>No. of Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quizzes</td>
<td>____%</td>
<td>___</td>
</tr>
<tr>
<td>Major tests</td>
<td>____%</td>
<td>___</td>
</tr>
<tr>
<td>Homework</td>
<td>____%</td>
<td>___</td>
</tr>
</tbody>
</table>
Portfolios ___% ___
Performance tests ___% ___
Classwork ___% ___
Note/Lab books ___% ___
Other ___% ___

10. What method have I chosen for combining the components to form a final grade?
- 100 point scale
- Percentage of total point system
- Other (specify) ____________________________

11. What method have I chosen for assigning final grades?
- Fixed percentage method
- Total point method
- Other (specify) ____________________________

12. How will I handle borderline cases? ____________________________
_______________________________________________________________
_______________________________________________________________
_______________________________________________________________

Summary
The more you teach and discuss grades with learners and their parents, the more your grading practices will change. Expect to revise your approach to grading many times in the course of your teaching career. What should not change, however, is the systematic and thoughtful nature of how you go about developing a grading plan. If you follow the steps outlined in this chapter, you may not have a perfect grading plan, but you will have one that makes sense to you and that you can explain and justify to learners and
parents.

Activities
1. Complete the Grading Plan Checklist at the end of the chapter, checking the appropriate boxes and providing the information requested, where appropriate.

2. Using the information you have provided on the Grading Plan Checklist, devise a final grade from combining scores for homework, objective tests, performance tests, portfolio, classwork, and notebook. Indicate whether you have chosen a 100 point scale or percentage of total point system for combining components and whether you have chosen the fixed percentage or total point method for assigning the final grade.

3. Prepare a written response to an angry parent justifying why you chose the weight you did on your Grading Plan Checklist for "objective tests" that measure recall and recognition and did not chose a larger weight for the portfolio and performance tests that measure higher thought processes.

4. A parent asks you at Parent-Teachers night to justify why her son got a "B" when he missed an "A" by only one point. She explains how hard her son has worked in your class and offers to have him do an assignment for extra credit, if you would change his grade to an "A." Prepare a response.

Suggested Reading

Concisely and clearly highlights the critical issues any teacher must grapple with when designing a grading plan.
Sample District Grading Policy

Assessments

• Should measure student knowledge and skills.
• Should be used to improve instruction.
• Should be valid, reliable, practical, and credible.

Communication

• Standards and criteria for student performance should be communicated in advance and actual student performance reported in detail following assessment.
• The reporting system should be clear and user friendly. All parties interested in the grading should know what the grade means and represents.
• The elements of the reporting system—grades, scores, and comments—are the primary means of communication from the teacher to the students and parents.
• Communication of scores and grades needs to be regular and frequent enough so as to be motivating and encouraging.

The Purpose Of And What Grades Represent

• The purpose of grades is to report student achievement related to clearly defined content standards.
• There will be separate reports on growth, effort, group work, participation, etc.
• The grades will be consistent from class to class with respect to what the grade is measuring.
Reporting System

- A reporting system should include, in addition to grades:
  progress/improvement
  effort
  attitude factors
  relative achievement
- A reporting system should be:
  fair
  accurate
  consistent
  honest

Specific Guidelines

- The course grade is based on student achievement.
- The course grade will be reported in an A-F format.
- No single assessment can count more than 20% of a course grade.
- Student performance on key performance factors will be reported to students and parents separately using an ‘O’utstanding, ‘S’atisfactory, and ‘U’nsatisfactory scale.
- These performance factors will be identified and communicated to students and parents at the beginning of each grading period.
- These factors may include the following: attendance, completion of work, cooperative behavior, effort, quality work, self directed learning, collaborative work, community contributions, complex thinking, and quality production.
• Academic grades will not be affected because of performance on developmental or practice tasks, homework used as practice, or performance in other areas, such as discipline, attitude, and/or attendance.

Figure 11.1. Sample District Grading Policy
<table>
<thead>
<tr>
<th>Grade</th>
<th>Absolute Scale</th>
<th>Relative Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Test Scores Average 90%</td>
<td>Top 10% of class</td>
</tr>
<tr>
<td></td>
<td>Papers, projects, etc. average to an &quot;A&quot; grade</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Command of knowledge is extensive</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Test Scores Average 80%</td>
<td>Top 33% of class</td>
</tr>
<tr>
<td></td>
<td>Papers, projects, etc. average to &quot;B&quot; grade</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Command of knowledge was well beyond minimum</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Test scores average 70%</td>
<td>Middle 20% of class</td>
</tr>
<tr>
<td></td>
<td>Papers, projects, etc. average to &quot;C&quot; grade</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Has command of basic knowledge</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Test scores average 60%</td>
<td>Bottom 33% of class</td>
</tr>
<tr>
<td></td>
<td>Papers, projects, etc. average to &quot;D&quot; grade</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lacks some basic knowledge</td>
<td></td>
</tr>
</tbody>
</table>
F

Test scores average below 60%
Papers, projects, etc. average below "D" grade
Has not learned most basic concepts and principles

Bottom 10% of class

Figure 11.2. Absolute and Relative Grading Plans
1. How active was ____ in helping the group attain its final product?

   ____ very active
   ____ fairly active
   ____ somewhat active
   ____ not too active
   ____ not active at all

2. How complete (or accurate, or useful, or original) is this groups final product?

   ____ very complete
   ____ fairly complete
   ____ somewhat complete
   ____ not too complete
   ____ not complete at all

Figure 11.3. Sample Scales for Evaluating Individual and Group Effort in a Collaborative Activity.
<table>
<thead>
<tr>
<th>Component</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework</td>
<td>15%</td>
</tr>
<tr>
<td>Objective tests</td>
<td>20%</td>
</tr>
<tr>
<td>Performance tests</td>
<td>25%</td>
</tr>
<tr>
<td>Portfolio</td>
<td>20%</td>
</tr>
<tr>
<td>Classwork</td>
<td>10%</td>
</tr>
<tr>
<td>Notebook</td>
<td>10%</td>
</tr>
</tbody>
</table>

Figure 11.4. Grading Components for an 8th Grade Life Science Class
<table>
<thead>
<tr>
<th>Component</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework</td>
<td>15%</td>
</tr>
<tr>
<td>Objective tests</td>
<td>20%</td>
</tr>
<tr>
<td>Performance tests</td>
<td>25%</td>
</tr>
<tr>
<td>Portfolio</td>
<td>20%</td>
</tr>
<tr>
<td>Classwork</td>
<td>10%</td>
</tr>
<tr>
<td>Notebook</td>
<td>10%</td>
</tr>
</tbody>
</table>

Homework Scores: 85, 79, 90, 80, 70; Average = 80.8 X 0.15 = 12.1 pts.

Objective Test Scores: 85, 77, 93; Average = 85 X 0.20 = 17 pts.

Performance Test Scores: 90, 100, 60, 80; Average = 82.5 X 0.25 (weight) = 20.6 pts.

Portfolio Scores: 75, 90; Average = 82.5 X 0.20 = 16.5 pts

Classwork Scores: 78, 82, 90; Average = 83.3 X 0.10 = 8.3 pts.

Notebook Scores: 86, 90, 80; Average = 85.3 X 0.10 = 8.6 pts.

Final Grade = 12.1 + 17 + 20.6 + 16.5 + 8.3 + 8.6 = 83.1

Figure 11.5. Combining Components into a Final Grade
REFERENCES


Baker, E. (1994). TO BE ADDED


Dweck (1975) TO BE ADDED.


Eccles, J. S. (1989). Bringing young women to math and science. In M. Crawford & M. Gentry (Eds.), Gender and thought: Psychological perspectives (pp. 36-58). New York: Springer Verlag


Fincham, Hokoda, & Sanders (1989). TO BE ADDED


Griffin, S. & Case, R. (1975). Rightstart: An early intervention program for insuring that children's first formal learning of arithmetic is grounded in their intuitive knowledge of number (Year 2 report submitted to the James S. McDonnel Foundation)


Hill (1984) TO BE ADDED

Hoy & Gregg (1994). TO BE ADDED

Hunt (1995). TO BE ADDED


National Educational Goals Panel (1993). TO BE ADDED


Resnick (1976). TO BE ADDED

Resnick, Siegel & Kresh (1971). TO BE ADDED


White (1974). TO BE ADDED