Pediatric Acute Care

A Guide for Interprofessional Practice

Edited by

Karin Reuter-Rice, PhD, NP, FCCM

Assistant Professor

School of Nursing
School of Medicine, Department of Pediatrics
Duke University
Durham, North Carolina

Beth Nachtsheim Bolick, DNP, NP

Associate Professor

Department of Women, Children, & Family Nursing
College of Nursing
Rush University
Chicago, Illinois

JONES & BARTLETT
LEARNING

World Headquarters
Jones & Bartlett Learning
5 Wall Street
Burlington, MA 01803
978-443-5000
info@jblearning.com
www.jblearning.com

Jones & Bartlett Learning
Canada
6339 Ormindale Way
Mississauga, Ontario L5V 1J2
Canada

Jones & Bartlett Learning
International
Barb House, Barb Mews
London W6 7PA
United Kingdom

Jones & Bartlett Learning books and products are available through most bookstores and online booksellers. To contact Jones & Bartlett Learning directly, call 800-832-0034, fax 978-443-8000, or visit our website, www.jblearning.com.

Substantial discounts on bulk quantities of Jones & Bartlett Learning publications are available to corporations, professional associations, and other qualified organizations. For details and specific discount information, contact the special sales department at Jones & Bartlett Learning via the above contact information or send an email to specialsales@jblearning.com.

The authors, editors, and publisher have made every effort to provide accurate information. However, they are not responsible for errors, omissions, or for any outcomes related to the use of the contents of this book and take no responsibility for the use of the products and procedures described. Treatments and side effects described in this book may not be applicable to all people; likewise, some people may require a dose or experience a side effect that is not described herein. Drugs and medical devices are discussed that may have limited availability controlled by the Food and Drug Administration (FDA) for use only in a research study or clinical trial. Research, clinical practice, and government regulations often change the accepted standard in this field. When consideration is being given to use of any drug in the clinical setting, the health care provider or reader is responsible for determining FDA status of the drug, reading the package insert, and reviewing prescribing information for the most up-to-date recommendations on dose, precautions, and contraindications, and determining the appropriate usage for the product. This is especially important in the case of drugs that are new or seldom used.

Production Credits
Publisher: Kevin Sullivan
Acquisitions Editor: Amanda Harvey
Editorial Assistant: Sara Bempkins
Production Editor: Amanda Clerkin
Associate Marketing Manager: Katie Hennessy
Permissions and Photo Researcher: Amy Mendosa
V.P., Manufacturing and Inventory Control: Therese Connell
Composition: diacriTech, Chennai, India
Cover Design: Scott Moden
Cover Image: © Prudkov/ShutterStock, Inc.
Printing and Binding: Courier Kendallville
Cover Printing: Courier Kendallville

Library of Congress Cataloging-in-Publication Data
Pediatric acute care : a guide for interprofessional practice/edited by Karin Reuter-Rice, Beth Bolick.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-7637-7971-9 (alk. paper)
1. Pediatric emergencies. 2. Pediatric intensive care. 3. Pediatric nursing. I. Reuter-Rice, Karin. II. Bolick, Beth.
[DNLM: 1. Critical Illness—nursing. 2. Acute Disease—nursing. 3. Child. 4. Critical Care—methods. 5. Infant. 6. Nurse Practitioners. WY 159]
RJ370.P382 2012
618.92'0025—dc22 2010040548

6048

Printed in the United States of America
15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

ONLINE CHAPTER 2

Rosemarie Suhayda and David Barnett

The Role of Assessment in Teaching and Learning

- **Purpose of Assessment**
- **Assessment Standards**
- **Levels of Assessment**
- **The Language of Assessment**
- **Formative and Summative Assessments**
- **Direct and Indirect Assessments**
- **Competency-Based Assessments**
- **Qualitative and Quantitative Assessments**
- **Developing Assessable Objectives**
- **Constructing Learning Objectives**
- **Domains of Learning**
- **Qualities of Effective Assessment and Performance Evaluation**
- **Using Rubrics to Assess Learning**
- **Summary**
- **References**

The education of health care professionals (HCP), particularly those in the acute care setting, is a dynamic process that has become increasingly complex and rigorous. Keeping abreast of the exponential increase in evidence-based and innovative practice is an educational challenge. Equally challenging is the assessment and measurement of the ever-expanding and complex skill set through which learners evidence their advanced knowledge, skills, and judgment. In the past, educational assessment was based on educational inputs and teaching activities. Today, however, the focus is on measuring learning outcomes. Decisions about modifications in curriculum and teaching–learning processes should be driven by analysis of how well learners achieve expectations that are based on predetermined learning outcomes.

PURPOSE OF ASSESSMENT

The major goals of educational assessment are to provide objective evidence that will help instructors make educational decisions about learners; to give feedback to learners about their progress, strengths, and weaknesses; to judge instructional effectiveness and curricular adequacy; and to inform policy. Why are these points important? In the health care setting, what and how learners learn affects the lives and well-being of health care consumers. Instructors have a professional and ethical responsibility to assess, evaluate, or graduate competent HCPs. To help less experienced instructors and individuals who do not fully understand the assessment process, McCarthy and Murphy (2008) recommend instructors make assessment strategies more transparent and user-friendly.

The focus of learning assessment is not on how well instructors teach, but rather on how well learners learn. Instructors spend a great deal of time developing responsive curricula and innovative teaching strategies. They select clinical experiences that will provide optimal learning. However, they may not take into consideration how well learners understand what is expected of them. High-quality assessments can provide instructors with the needed confidence that their learners are meeting those learning expectations.

ASSESSMENT STANDARDS

In response to increased public and professional awareness of the need for assessment competence, the American Federation of Teachers, National Council on Measurement in Education, and National Education Association developed *Standards for Teacher Competence in Educational Assessment of Students* (1990); meeting these criteria is critical to the role of educators (Table 2-1). As a collective entity, these standards call on instructors to demonstrate skill at selecting, developing, applying, using, communicating, and evaluating learner assessment information and practices.

TABLE 2-1

Standards for Teacher Competence in Educational Assessment of Students

1. Teachers should be skilled in choosing assessment methods appropriate for instructional decisions
2. Teachers should be skilled in developing assessment methods appropriate for instructional decisions
3. Teachers should be skilled in administering, scoring, and interpreting the results of both externally produced and teacher-produced assessment methods.
4. Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement.
5. Teachers should be skilled in developing valid pupil grading procedures that use pupil assessments.
6. Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.
7. Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.

Source: American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *Standards for teacher competence in educational assessment of students*. Washington, DC: National Council on Measurement in Education.

LEVELS OF ASSESSMENT

Miller and Leskes (2005) identify four levels at which learning assessment should take place: learner, course, program, and institution.

Within individual courses, assessment examines how well learners meet the course objectives. The presumption is that learners have a clear understanding of what they need to learn and how their learning will be assessed. Data used for assessing learning within a course include all assignments embedded within the course. Data are gathered sequentially, and feedback is typically given to learners in the form of instructor-assigned grades or written feedback on clinical performance. Learners use instructor feedback to track their own performances within a course and to identify areas in need of improvement. Instructors use data to evaluate their effectiveness in communicating information and in motivating learners to exert efforts needed to achieve quality work.

Assessing individual learning across courses helps to establish how well learners learn throughout the span of their programs. Data sources used for evaluating learning at this level include portfolios consisting of samples of leaner coursework and instructor feedback assembled from

a number of designated courses, capstone experiences or projects, logs or journals, and externally developed standardized exams. Learners use these assessments to evaluate their own progress throughout a program. Instructors use these data to track learner progress through a program of study.

Assessment of courses is based on the collective achievements of learners within a particular course. Data sources used to evaluate courses include course portfolios constructed by the course instructors. Portfolios might include the following items: course syllabi or outlines and course expectations; descriptions of course assignments and their alignment with the course objectives; test blueprints; samples of the leaner's work; examinations, both externally and internally developed; clinical evaluation forms, where applicable; communication with learners; peer evaluations; and course evaluations. Aggregate course data obtained from learners within a course can be used to evaluate how well course content and assignments achieved the designated learning outcomes. Curriculum committees and instructors can use course assessment data to plan future improvements in a course.

Program assessment helps to establish how well the collection of courses and learning experiences helps to fulfill program-wide goals and expected learning outcomes. Assessment at this level should address how the sequencing of courses and curriculum design allow for cumulative learning across the program. The alignment between professional standards, program outcomes, and course outcomes should be assessed as well. Data sources for program-level assessments are similar to those already mentioned, and data can be gathered at multiple points (entry into the program, midpoint of the program, end of the program). Particularly valuable data include summative indicators of performance such as standardized tests, capstone experiences, and final papers that demonstrate a synthesis of knowledge and skills.

Institutional-level assessment is performed primarily to fulfill criteria for institutional accreditation and to demonstrate to external stakeholders the value and quality of the educational programs. A major objective of this type of assessment is to provide credible data that informs planning for internal improvement. At this level, the institution demonstrates how well each educational program contributes to institutional outcomes. It presumes alignment between institutional goals and mission and learning outcomes for each of the educational programs.

THE LANGUAGE OF ASSESSMENT

Several common, yet often confusing, terms are associated with the assessment process. Three such terms are *assessment*, *measurement*, and *evaluation*. Although they are frequently used interchangeably, these terms are not equivalent. Assessment refers to the gathering of learner information, whereas evaluation involves a judgment about data obtained through measures or measurement procedures. Measurement is the process of ascribing a number or score to a characteristic, trait, or behavior. For example, instructors observe the clinical performance of learners (assessment), rate their performance by using grading rubrics (measure), and judge whether the performance met pre-established standards (evaluation).

Evaluation assumes that the observed behavior or performance is compared with a pre-established standard. At course and program levels, standards are most often depicted as learning objectives or program outcomes, but may also take the form of grading or progression criteria. In addition, instructors use professional standards to judge the competency of their learners, such as those developed by professional organizations (e.g., the American Heart Association and the American Academy of Pediatrics for use in the Pediatric Advanced Life Support [PALS] Provider Course). Regardless of the discipline, learners should be made aware of those standards by which their performance is being judged before instruction and evaluation begin. Clearly establishing and communicating assessment and evaluation criteria serves multiple purposes: It facilitates the evaluation process, it guides learners in determining what is important to learn, and it helps learners appreciate how to self-assess themselves.

Effective assessments and evaluations continuously yield information on how well learners are achieving learning goals. They are outcome oriented, evidence based, quality focused, and understood and valued by instructors and learners alike. In the ideal assessment model, learning goals are clearly aligned with the program mission and goals, learning supersedes coverage of content, applying knowledge supersedes knowing, improving learning supersedes reporting assessment data, and conversations about improvement are commonplace.

FORMATIVE AND SUMMATIVE ASSESSMENTS

Assessment is an iterative, participatory activity involving learners and instructors. *Formative assessment* is a diagnostic process that evaluates learning during experiences and appraises learner progress toward meeting course and/or program objectives. Formative feedback is used by learners to help them identify their strengths and weaknesses; it is used by instructors to evaluate teaching effectiveness.

Chappuis (2005, pp. 40–42) suggests several strategies to help promote learner involvement in the formative assessment process.

1. *Provide a clear and understandable vision of the learning target*. For formative assessment to be of value, learners need to know the learning targets or standards against

which they are being evaluated. This strategy includes sharing the scoring guide or rubrics used to evaluate the learner. Notably, formative assessment aims to shape the learning process, it should *not* be tied to a grade.

2. *Use examples of strong and weak work*. For comparison, learners need to know what constitutes excellent, satisfactory, or unsatisfactory performance. Chappuis finds it useful to have learners apply the scoring rubrics to the work of others and then to defend their judgments. This exercise also helps them develop skills in self-assessment.
3. *Offer regular descriptive feedback*. Feedback, to be useful, must have meaning. Telling learners that they are doing well or not well without descriptors of the specific behaviors provides little guidance for either eliminating unsatisfactory behaviors or replicating good ones. An example of descriptive feedback might be, "Your assessment techniques are very organized but you will need to ask more follow-up questions to obtain more complete information."
4. *Teach learners to self-assess and set goals.* Instructors should model not only clinical behaviors, but also evaluation techniques. Giving learners' feedback that is useful for their self-improvement provides them with standards against which they can judge their own performance. Encouraging them to use the evaluation rubrics also helps them develop their skills of self-assessment.
5. *Design learning experiences to focus on one aspect of quality at a time*. Referring to the earlier example, if a learner is having difficulty with a particular aspect of assessment and the scoring rubric focuses on assessment skills, then the learning experience should center on the aspect with which the learner is having difficulty. The learner should be encouraged to use the scoring rubric to assess his or her skill level. It is important to allow learners sufficient time to use feedback and to practice and improve their performance before a grade is assigned.
6. *Engage learners in self-reflection and let them document and share their learning.* Encourage learners to monitor their progress and improvement over time. The documentation of this ongoing process may take the form of logs or compilations of work, for example.

Summative assessment is a comprehensive appraisal of learners learning that occurs upon completion of the instructional activity and typically culminates in a grade. Like formative assessment, summative assessments should be consistent with the course and program learning objectives and should be clearly communicated to learners. Examples of summative assessments include unit, comprehensive, or final examinations, either standardized or locally developed; final papers; presentations (as in capstone requirements); demonstration/psychomotor skills or simulation performances; and community-, civic-, or organizational-level projects.

Although grades are typically associated with summative assessments, grades in and of themselves are not considered good measures of learning. Furthermore, most accrediting bodies do not recognize grades to be sufficient evidence when they are presented as the sole proof of learning. For example, at the course level, an instructor might report that 80% of learners achieved 90% or higher on an examination. Without further analysis, the score gives little information about actual learning. In another example, a learner might receive an A for a paper or presentation; if an analysis of the paper's or presentation's strengths and weaknesses is not provided, however, the learner has little information to guide improvement. The same is also true for clinical evaluations: Learners might receive grades of "pass" or "satisfactory," yet have little understanding of areas in which they excel or perhaps require more work. At the program level, a certification pass-rate gives evidence of program performance relative to a national group comparison or pre-established standard, but provides little guidance to instructors on program improvement. If available, further analysis of the content or topic areas on which learners are judged would provide more useful data for curriculum guidance and program improvement.

DIRECT AND INDIRECT ASSESSMENTS

Assessments can be either direct or indirect. *Direct assessment* constitutes tangible, visible, self-explanatory evidence of exactly what learners have and have not learned. Examples of direct assessments include clinical simulation, demonstration/psychomotor skills, or practicum performance; achievement on course or standardized tests; and work submitted in the form of portfolios, writing assignments, and capstones. To be most effective, direct measures should be aligned with specific learning objectives. *Indirect assessment* constitutes performance quality deduced from secondary sources, such as learner logs, as well as surveys administered to learners, alumni, and employers. The goal in assessment is for direct data sources to outweigh indirect sources.

COMPETENCY-BASED ASSESSMENTS

The terms *outcomes, characteristics, skills, abilities, knowledge,* and *competencies* are often used interchangeably to describe the results of the learning process. To provide some clarity, the National Postsecondary Education Cooperative (NPEC), under the U.S. Department of Education National Center for Education Statistics (2002), developed a hierarchical differentiation of terms (Figure 2-1).

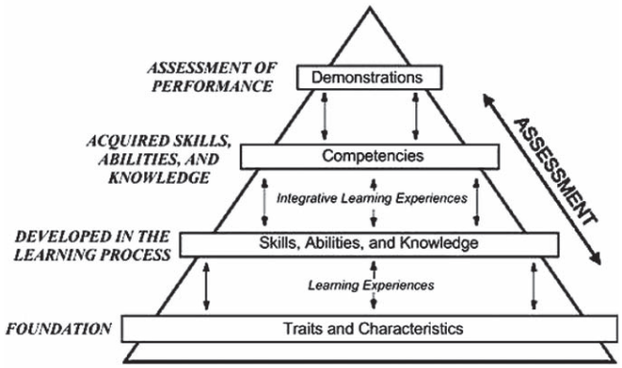

FIGURE 2-1

Hierarchy of Postsecondary Outcomes.

Source: U.S. Department of Education, National Center for Education Statistics. *Defining and Assessing Learning: Exploring Competency-Based Initiatives*, NCES 2002–159, prepared by Elizabeth A. Jones and Richard A. Voorhees, with Karen Paulson, for the Council of the national Postsecondary Education Cooperative Working Group on Competency-Based Initiatives. Washington, DC.

In this model, traits and characteristics serve as the foundation for learning and consist of the innate attributes of individuals on which further experiences can be built. Skills, abilities, and knowledge are developed through learning experiences. Competencies are the results of integrative learning experiences and represent the combination of skills, abilities, and knowledge. Demonstrations are the results of applying competencies and constitute the level at which performance can be assessed. The challenges for instructors are to identify and describe the optimal bundle of skills, abilities, and knowledge that constitute a competency; to select multiple means for measuring the competency; to create and ensure learning experiences that lead to the competency; and to identify a standard against which the learner is judged to be competent. Table 2-2 gives an example of how competencies, learning experiences, and demonstrations of learning can be linked.

According to the NPEC, it is the bundling of skills, abilities, and knowledge that defines competencies needed in a given situation; furthermore, competencies may change depending on the context or situation in which they are applied. For example, assessment in a geographical community or population may require a different set of competencies than does assessment in an acute care setting. The NPEC (2002) also developed several principles to guide instructors in the process of competency-based initiatives and assessment:

1. The appropriate stakeholders fully participate in identifying, defining, and reaching a consensus about important competencies.
2. Competencies are defined at a sufficient level of specificity that they can be assessed.
3. Multiple assessments of competencies provide useful and meaningful information relevant to decision making.
4. Assessments of competencies are directly linked with goals of the learning experience.
5. Assessment results are used in making critical decisions about strategies to improve learner learning or allocation of resources.

The American Nurses Association, in its most recent draft of its *Scope and Standards of Practice* (2010), also provides definitions of skills, abilities, knowledge, and judgment. According to this organization, skills include

TABLE 2-2

Assessment of Learning		
Competency Description	**Learning Experiences**	**Demonstrations of Learning**
Relates physiologic, psychosocial, and epidemiologic principles and theories to the care of the pediatric patient population.	Answers essay questions and gives oral and written case presentations.	The student correctly distinguishes between normal and abnormal developmental and age-related physiologic changes associated with the pediatric patient population. The student provides a correct scientific rationale for patient management decisions. The student correctly identifies risk, patterns of disease, and effectiveness of prevention strategies in the pediatric patient population.
Functions as a teacher-coach with pediatric patients and their families.	Designs and implements a learning plan tailored to patient/family needs.	The student correctly assesses patient/family readiness and ability to learn. The student elicits the patient's/family's perception of health, barriers/supports to learning, learning style, and cultural influences on learning.

psychomotor, communication, interpersonal, and diagnostic skills. Knowledge encompasses thinking about and understanding of science, humanities, and professional standards of practice, as well as insights gained through experience. Ability is the capacity to act effectively. Judgment involves critical thinking and problem solving along with ethical reasoning and decision-making abilities.

QUALITATIVE AND QUANTITATIVE ASSESSMENTS

Learning assessments can be obtained either through qualitative or quantitative avenues.

Qualitative assessment emphasizes a holistic approach that looks at those complex elements of learner performances that cannot be easily measured or quantified. The resulting assessments are subjective in nature and are best summarized as descriptive statements from observations, portfolio reviews, interview transcripts, videotaped performances, and written comments on questionnaires or in journals (McCready, 2007). Qualitative approaches typically take less time in the planning phase and more time in the analysis phase. Most preliminary effort will be spent developing the protocol used to collect the data.

Quantitative assessments are more objective in nature and come from a numerical-measurement model. Performance evaluations are based on numerical data, such as frequencies, averages, and proportions obtained from tests or numerical scales, and can be analyzed with statistical procedures. For quantitative assessments, the work effort in designing the assessment occurs upfront in the creation of the assessment tool. After that point, the analysis is relatively straightforward. Typically, quantitative assessments provide data on learner rankings relative to a standard, which allows for comparisons within or across cohorts of learners. Quantitative assessments that are carefully designed using multiple-choice questions based on clinical scenarios, simulated patient care situations, and case studies can also be effective teaching tools aimed at helping learners develop and apply their critical thinking knowledge and skills (Leung et al., 2008).

Qualitative assessment can be more congruent with program objectives that value developing lifelong learners and learner-centered strategies, although not all qualitative approaches work equally well with all learners. For example, Chirema (2007), in a study examining the use of journaling to promote reflection and learning, found that only two-thirds of the learners were able to demonstrate varying levels of reflection through the journaling process, rendering journaling an ineffective approach to assessment. Quantitative approaches present a greater challenge for measuring concepts such as independence, self-initiative, ability to work with a team, ability to manage a crisis, and so on. These characteristics can be rated using scales; however, scales tend to reduce a complex set of skills to a few scale points. In doing so, the richness of a learner's experience may be lost.

DEVELOPING ASSESSABLE OBJECTIVES

Objectives play an important role in assessment because they define the acquisition of knowledge, skills, attitudes, and values that represent the outcomes of learning. Terms used interchangeably with "learning objectives" include "learning outcomes," "competencies," and "terminal objectives." The term "goal" is often confused with the term "objective." Goals are statements that, in broad terms, describe what is expected of the learner, whereas objectives are statements of specific and measurable behaviors that describe what the learner will know and be able to do as a result of the educational experience.

Guided by the program mission, instructors devise behavioral statements reflecting competencies expected of learners as they progress through an educational program. These statements serve as the basis for developing assessment strategies. The emphasis is not on the educational process or teaching methods, but instead on the educational product, which is learning. For example, consider the following two objectives:

- The instructor will demonstrate steps in completing a comprehensive physical examination.
- The learner will perform a comprehensive physical examination.

The first objective is instructor centered, whereas the second is learner centered. The broad learning objective of "performs a comprehensive physical examination" can be further divided into more specific behaviors or outcomes. In this case, that set of elements would include the components of a comprehensive physical examination.

For assessment to be effective, learning objectives must be explicit enough to ensure both learners and instructors understand the criteria being used to judge learner performance. Burns and his colleagues at Central Michigan University (n.d.) suggest four steps be followed in developing learning objectives:

1. Examine the department, college, and university mission statements.
2. Determine what graduates of your program should know, which skills they should be able to demonstrate, and which professional values they should hold.
3. Convert the list of expected knowledge and skills into general program objectives.

4. Convert the general program objectives into specific statements that specify the knowledge and skills expected of learners at the course level.

CONSTRUCTING LEARNING OBJECTIVES

Robert Mager (1984), in his classic text *Preparing Instructional Objectives*, divides learning objectives into three subcomponents:

- The behavior that should be demonstrated—that is, what the learner will be able to do as a result of the learning activity
- The condition under which the behavior is to be completed—that is, how the learner will be able to demonstrate the behavior
- The standard or level of performance against which the behavior will be judged

In contemporary education, particularly in clinical practice, the conditions and level of performance are frequently omitted from the learning objective. It is assumed that these conditions involve the practice setting and the standards are set at a very high level of performance. Nonetheless, the expected learning behavior should be expressed in specific, measurable, observable terms.

Well-written learning objectives begin with an action verb such as "define," "diagram," "explain," "apply," "analyze," "formulate," "compare," "contrast," or "appraise." Vague words such as "understand," "appreciate," and "know" should be avoided. Without further descriptors, these terms are not measurable. Also, each objective should define only one behavior. For example, the objective "Describe the clinical manifestations of pain in an infant and devise a pharmaceutical pain management plan" includes two behaviors within a single statement. A more appropriate approach is to state each behavior separately: "Describe clinical manifestations of pain in an infant" and "Develop a pharmaceutical pain management plan for an infant." Note that related behaviors can be combined into a single objective—for example, "Develop a pain management plan that includes both pharmacologic and nonpharmacologic approaches."

DOMAINS OF LEARNING

Learning domains can be categorized as cognitive, affective, and psychomotor. It is recommended that learning objectives focus on a single domain; statements that combine domains should be avoided.

COGNITIVE DOMAIN

Objectives written within the cognitive domain focus on the learner's knowledge, application, and synthesis of the subject matter. Lowest levels of learning emphasize recall and include knowledge and comprehension. Knowledge is rote memorization, and comprehension is the restatement of memorized facts (Morrison et al., 1996). Action verbs that represent knowledge and comprehension include the following: define, label, list, discuss, classify, outline, paraphrase, report, and summarize. Application and analysis represent interpretation of information and are characterized by action verbs such as these: adapt, apply, compare, contrast, criticize, predict, develop, employ, and illustrate. The highest levels of learning encompass synthesis and evaluation and reflect problem-solving abilities. Action verbs at this level include the following: assess, conclude, conceive, formulate, diagnose, evaluate, judge, and recommend. The cognitive domain is often evaluated with written examinations, papers, and oral presentations or explanations.

Instructors often utilize Bloom's taxonomy of the cognitive domain when constructing course learning objectives (Anderson et al., 2001). More than 50 years ago, Bloom (1956) reported that more than 95% of the examination questions he reviewed required only demonstration of low-level thinking skills such as recall or recognition. Unfortunately, some contemporary classroom examinations fare little better. To assess highest-level thinking processes, Morrison et al. (1996) encourage assessing multilogical thinking—that is, application and analysis of two or more facts or concepts within the same evaluation measure. Educational research indicates that learners who learn material and then demonstrate higher-level learning through application, synthesis, and evaluation remember more of what they have learned (Scott, 2003; Zheng et al., 2008).

AFFECTIVE DOMAIN

The affective domain encompasses learner motivation, attitudes, perceptions, and values. Examples of objectives within the affective domain are "Accepts responsibility for his or her actions" and "Respects the beliefs of others." Behaviors within the affective domain are difficult to measure, because they are less observable and more subjective in nature. Furthermore, it would be difficult to causally link these behaviors with learning activities. For these reasons, *it is recommended that use of objectives addressing the affective domain be limited.*

PSYCHOMOTOR DOMAIN

The psychomotor domain is skill based. Thus objectives within this domain focus on specific skills that involve physical movement, coordination, and use of motor skills. Measures of psychomotor competencies involve imitation, precision, manipulation, and techniques of execution. The following action verbs are frequently associated with the psychomotor domain: demonstrates, conducts, displays, exhibits, and performs.

QUALITIES OF EFFECTIVE ASSESSMENT AND PERFORMANCE EVALUATION

A key element in any assessment process is establishing the validity and reliability of the methods and instruments used to evaluate performance. *Validity* is the extent to which an instrument measures what it purports to measure. *Reliability* is the extent to which an instrument allows consistent measures of the performance.

Validity and reliability are independent constructs. It is possible, for example, to have an assessment with low reliability and high validity; this result indicates that the test obtains good information, but does so inconsistently. It is also possible to have an assessment with high reliability and low validity, which indicates that the test consistently obtains bad information. Because the results of assessment affect decisions about learner progression, it is important that measures used to evaluate learner learning and performance are both accurate and trustworthy.

VALIDITY

The most commonly employed approach when looking at validity is the c*lassical* or *conventional* model, which includes face, content, criterion-related, and construct validity.

Face validity

A test or assessment method is said to have face validity if it "looks like" it measures what it is supposed to measure. Face validity is exceedingly vague and is considered the weakest form of validity. Psychometricians have not used this form of validity for decades and do not recommend it as a sole criterion for selecting assessment instruments. With face validity, curricular decisions may be made based on a common-sense approach as opposed to evidence-based approaches that show actual learner outcomes may be counter-intuitive to what one might expect. Nonetheless, some educators continue to ascribe face validity to assessment methods based on an instrument's ability to demonstrate a "common sense" or "persuasive" explanation for assessment findings (Lacity & Jansen, 1994).

Content validity

Content validity represents the extent to which a measurement reflects the specific intended content domain or the extent to which it represents all facets of a concept. With content validity, instructors can infer that the items on a given test or measure represent the larger domain of knowledge and skills that the learner is expected to learn. One approach to establishing content validity is to map the knowledge and skills being evaluated to the learning objectives of the course or program. In evaluation of clinical performances, for example, educational expectations may be mapped to the standards or learning objectives that frame the experience. In test construction, this process is referred to as "test blueprinting."

A test blueprint helps instructors establish the content validity of their tests by ensuring that all of the major learning objectives are addressed by the testing instrument. Blueprints also help ensure that items with varying degrees of cognitive difficulty are included. Because classroom tests are often the method of choice for evaluating learner achievement, a test blueprint serves as the foundation or architectural structure for test development.

Test blueprints consist of three components displayed in a two-dimensional matrix or chart. The components include the content or learning objectives to be measured, the thinking or cognitive processes to be tested, and the relative weight ascribed to both the content and cognitive areas.

To construct the blueprint, list the learning objectives or content areas to be measured along the left column of the matrix. Next, list the cognitive levels along the top of the matrix. Ideally, the majority of items will be written at the higher levels of cognition. Within each cell of the matrix, indicate the proportion of items that represent the intersection between the content or learning objectives and the cognitive level.

Consider the blueprint in Table 2-3, which applies to a test that will have 40 items and will cover three content areas. Cognitive levels are compressed into three categories: knowledge/comprehension, application/analysis, and synthesis/evaluation. Instructors decide that 25% of the items will focus on illness assessment, 50% on illness management, and 25% on illness prevention. Next, instructors establish a distribution of items within each content area. Distribution should be determined by level of importance, not by the amount of time spent on a particular content area.

Test blueprints can be constructed either before or after test items are written. If a test bank already exists, start by assigning a number or identifier to each item in the test bank, and then categorizing each item according to its cognitive level and content area. The item number or identifier can then be placed in the matrix cells once a decision has been made about how items will be

TABLE 2-3

Test Blueprint

Content	Cognitive Levels			Total Number of Items
	Knowledge/ Comprehension	Application/ Analysis	Synthesis/ Evaluation	
Illness assessment	1 (10%)	4 (40%)	5 (50%)	10 (25%)
Illness management	2 (10%)	10 (50%)	8 (40%)	20 (50%)
Illness prevention		4 (45%)	6 (55%)	10 (25%)
Total number of items	3 (8%)	18 (45%)	19 (48%)	40 (100%)

distributed. For example, instructors might select test item number 10 to be included as one of the four items used to test application/analysis in the area of illness prevention.

Criterion-related validity

Criterion-related validity refers to assessments that predict concurrent or future performance. In health care education, this typically occurs in a clinical practicum or simulated exercise, which mimics the working environment of advanced HCPs. Concurrent validity is a form of criterion-related validity that measures how well a particular measure correlates with a previously validated measure administered at approximately the same time. Consider, for example, first administering a previously validated standardized examination to a group of learners, and then administering a locally developed (instructor-developed) examination that purports to test the same content. A high correlation between the two tests indicates that they have concurrent validity. One could infer that the locally developed examination is a valid measure of the same content domain as is the standardized examination.

Predictive validity measures how well an instrument will predict future performance. Consider again the standardized or locally developed test in relation to the national certification examination. A high correlation between these tests implies that learner performance on the locally developed or standardized test will predict performance on the certification test. Predictive validity can also be applied to the assessment of clinical performance. For example, a learner's performance on a valid stimulation exercise can be compared with performance in the clinical area. If the simulation and clinical performance are comparable, then it can be inferred that the simulation exercise has predictive validity.

All types of validity require interpretation of the data obtained through the instrument. Evidence may be organized in several ways to support the validity of an interpretation. For example, evidence can be gathered on the following aspects of performance:

- Test content
- Response processes
- Internal structure of the assessment
- Relations to other variables that may affect performance
- Consequences of testing (e.g., is it high stakes?)

Instructors should carefully consider which of these kinds of evidence will prove useful in either reinforcing or revising propositions required for an interpretation being considered. Once collected, these types of evidence can be integrated into a validity argument. Such an argument may indicate that an assessment measure needs to be revised. that instructors need to adjust how the data are gathered, or that instructors need to reconsider the theoretical constructs underlying the interpretations of the data.

Construct validity

Construct validity seeks to establish agreement between an unobservable characteristic or trait. Construct validity incorporates two subcomponents: convergence and divergence. The principle underlying the determination of construct validity is that measures of two constructs that should be related are in fact related, as determined by high correlation coefficients, whereas measures of constructs that are opposite or divergent should not be related, as determined by low correlation coefficients.

Holmboe et al. (2003) described an example of construct validity as it applies to a clinical assessment tool. Forty instructors used a nine-point MiniClinical evaluation exercise (miniCEX) to rate the performance of standardized medical residents on nine scripted clinical videotapes. Each videotape depicted three levels of performance: unsatisfactory, marginal/satisfactory, and high satisfactory/superior. The rating scale also corresponded to the three levels of performance. Instructors rated performances for three clinical skills—namely, history taking,

physical examination, and counseling. Construct validity was established when instructors were able to successfully discriminate between the three prespecified performance levels using the miniCEX scale.

RELIABILITY

Reliability is the extent to which an assessment yields consistent results, given an unchanged measured phenomenon, over repeated measurements. According to Gronlund and Linn (1990), "Next to validity, reliability is the most important characteristic of evaluation results. Reliability (1) provides the consistency that makes validity possible and (2) indicates how much confidence we can place in our results" (p. 77). Reliability is estimated in several ways: test–retest reliability, internal consistency, and inter-rater reliability.

Test–retest reliability measures stability of an assessment over time and is the most straightforward approach to assessing reliability. To establish this type of reliability, the same instrument or test is administered to the same subjects at two points in time. A typical interval for a test–retest procedure is several weeks, although the appropriate length of time depends on how vulnerable the measure is to intervening factors or conditions. For example, a short interval between administrations could yield a falsely elevated correlation, mostly attributed to recall. Learners might learn from the first test and adjust answers accordingly on the next administration. Nonetheless, test–retest is still a widely used procedure, and a report by McKelvie (1992) indicated that memory effects may not always influence reliability estimates under a test–retest design.

Test–retest reliability is calculated as the correlation between scores from the two assessment periods, with 1.0 considered a perfect correlation. Typically, a correlation coefficient in the range of 0.70 to 1.0 is considered satisfactory or strong, and a correlation coefficient less than 0.30 is considered unsatisfactory or weak.

Internal consistency measures establish the correlation between different items on the same test. The most common types of internal consistency measures include *Cronbach's alpha,* which establishes correlations among all items on a test, and *Spearman-Brown split-half* reliability, which establishes correlations based on a test that is split into two halves. By convention, an item with a Cronbach's alpha value greater than 0.70 is considered adequate and one with a value greater than 0.80 is considered good. When using the Spearman-Brown split-half reliability coefficient, a coefficient greater than 0.80 is considered adequate and one greater than 0.90 is consider good. For both the alpha and split-half measures, a coefficient of 0.60 is considered a lenient cut-off value; a coefficient less than 0.50 is considered to indicate poor reliability. The *Kuder-Richardson coefficient* (KR20) is the same as Cronbach's alpha when applied to dichotomous items.

Inter-rater reliability is a measure of homogeneity in results when two or more raters assess the same subjects. Two procedures are used to measure inter-rater reliability: *Cohen's kappa,* which is applied when two raters are used, and intraclass correlation (ICC), which is employed when multiple raters are used. A Cohen's kappa value in the range of 0.40 to 0.59 is considered to indicate moderate inter-rater reliability, a value in the range of 0.60 to 0.70 to indicate substantial inter-rater reliability, and a value of 0.80 to indicate outstanding inter-rater reliability. ICC values are interpreted similarly to Cohen's kappa.

USING RUBRICS TO ASSESS LEARNING

Rubrics are descriptive, scoring guidelines or criteria used to rate performance, products, or processes of learning. Their purpose is to clearly define the criteria by which work will be judged. Rubrics also help maintain consistency for judging work when multiple instructors are involved in the evaluation process, and they guide instructors in giving both descriptive and evaluative feedback. Learners can use rubrics to improve their self-assessment skills as well as to provide feedback on peer-evaluated assignments. Although applicable to nearly all types of learning activities, rubrics are most often applied to academic and clinical competencies, which involve producing scholarly work, critiquing and synthesizing information, and applying newly acquired principles and concepts.

Rubrics are commonly classified as *analytic* or *holistic*, depending on whether learner performance is evaluated on multiple attributes or more globally on an entire task, respectively.

ANALYTIC RUBRICS

With analytic rubrics, performance is divided into two or more criteria or elements, each of which is evaluated on a separate scale. The criteria and scale are typically displayed in a matrix format, with performance criteria listed along the left-hand column of the matrix and scoring scale across the top. Performance criteria consist of the behaviors or learning objectives being evaluated. Scoring criteria or scales describe how well or poorly the component or objective is met and consist of numbers, letter grades, or labels. Most authors recommend the use of three to five scoring criteria and at least two performance criteria or elements within a matrix.

Rubrics typically include qualitative performance descriptors at each level of the scoring scale. Descriptors

TABLE 2-4

Evaluation Rubric for Completing a Patient History and Physical Examination

Performance Criteria	Excellent	Competent	Needs Work
Performs comprehensive health history and physical examinations	Consistently includes all relevant components of a comprehensive health history and correctly performs physical examination techniques	Completes the history and physical examination with only minor inaccuracies in content and technique	Routinely needs guidance in obtaining history data and performing physical examination techniques

TABLE 2-5

Evaluation Rubric for Clinical Practicum Objectives

1. Does not meet performance component
2. Needs continuous guidance to meet performance component
3. Demonstrates progress toward meeting the performance component
4. Consistently meets performance component, independently or with minimal guidance

Performance Criteria	Performance Level			
Performs comprehensive health history and physical examinations	1	2	3	4
Employs appropriate screening and diagnostic strategies	1	2	3	4
Develops pertinent differential diagnoses	1	2	3	4

FIGURE 2-2

Examples of Common Scoring Labels.

define more clearly which type of performance is expected at each scoring level. Table 2-4 and Table 2-5 present alternative examples illustrating how analytic rubrics with qualitative descriptors can be applied to the learner's ability to complete a health history and physical examination. Examples of common scoring labels appear in Figure 2-2.

Performance criteria can also be weighted and combined with numerical scoring criteria to assign a grade for the learning activity. Table 2-6 gives an example of a matrix that contains a weighted multiplier that can be used to calculate a score for the performance component. Scores can then be summed across criteria to determine the grade.

HOLISTIC RUBRICS

Holistic rubrics are used to evaluate performance criteria in combination, or holistically, on a single scale, as opposed to each criterion being evaluated separately. Holistic rubrics are particularly useful for written assignments, presentations, or group discussions. Table 2-7 gives an example of a holistic rubric used to assign a grade to the quality of learner participation in online discussions.

RECOMMENDATIONS FOR DEVELOPING RUBRICS

Moskal (2003) provides six recommendations that are pertinent to the development of either analytic or holistic rubrics:

1. Performance criteria included in the rubric should be clearly aligned with the stated course goals and objectives.
2. Performance criteria should be stated in terms of observable behaviors.
3. Performance criteria should be fair and unbiased, so that an unfair advantage is not given to a particular subset of learners.

TABLE 2-6

Weighed Performance Components for a Written Assignment

Performance Criterion	Weight	Poor (1)	Good (2)	Excellent (3)
Analyzes current research and clinical literature relevant to the key issues of the clinical practice problem.	× 5	Limited literature review. Limited analysis of research and its application to the clinical problem.	Literature review includes most relevant sources with accurate analysis. Identifies most issues associated with the clinical problem.	Comprehensive analysis of research and clinical literature. Analysis is thorough, appropriately related to the clinical problem, and integrated throughout the entire paper.

TABLE 2-7

Holistic Rubric for Posting Discussions in an Online Course

Grade	Scoring Criteria
A	Presents well-developed ideas with rationale; demonstrates application and synthesis of content; introduces new ideas; stimulates further discussion
B	Presents ideas that are developing with limited rationale; demonstrates some application and synthesis of content; sometimes stimulates discussion
C	Presents poorly developed ideas based mostly on opinion; does not add substantially to the discussion; demonstrates little or no application of content
F	Does not enter discussion; responds inappropriately or refers to inaccurate material

4. The number of points or values used in the scoring scale should reflect the importance of the activity or criterion.
5. The scale used for a scoring rubric should be distinct in differentiating between scoring levels.
6. Scoring criteria and rubrics should be clearly written in language that learners will understand.

VALIDITY AND RELIABILITY CONCERNS WITH RUBRIC DEVELOPMENT

Concerns about validity and reliability of the assessment rubrics are similar to those that arise when establishing any assessments and should be considered before the rubrics are developed and implemented.

Validity of assessment rubrics

The validity of rubrics is most appropriately categorized as content, construct, or criterion related.

Content-related validity, as it applies to rubrics, includes two components: (1) the extent to which the scoring rubric can be used to assess learner's knowledge of the content area being evaluated and (2) the extent to which the rubric samples the content domain (Moskal & Leydens, 2000). Referring back to Table 2-6, the specified performance criterion is "Analyzes current research and clinical literature relevant to the key issues of the clinical practice." The intended content of this task is analysis; therefore, the rubric associated with assessing the learner's ability to analyze research should encompass only those aspects associated with analysis. If, for example, the specific scoring rubric were to focus on assessment of the learner's writing abilities, then the inference about learner performance would relate to writing skill, not analytic skill, rendering the interpretation invalid relative to the content area being evaluated. Similarly, in Table 2-5, the performance criterion is "Performs comprehensive health history and physical examinations." In this instance, the content domain includes skills in both obtaining a health history and performing a physical examination. If the scoring rubric focused only on obtaining a health history, then the entire content domain is not being sampled.

Construct-related validity involves processes internal to the individual, such as reasoning abilities, psychological traits, personal qualities, and attitudes. Consider the holistic rubric in Table 2-7, which combines both content and construct validity within the holistic scoring criteria. "Presents well-developed ideas with rationale" refers to the learner's knowledge of the content. "Demonstrates application and synthesis of content" gives evidence of reasoning abilities.

Assessments with criterion-related validity should directly reflect the types of experiences encountered in real practice. Consider again the rubric in Table 2-5. The performance criteria listed pertain to the learner's ability to perform comprehensive health history and physical examinations. The scoring rubric, which ranges from 1 to 4, describes the type of performance expected at each level. High scores on these activities should suggest high performance levels in the work place.

Moskal and Leydens (2000, p. 5) identify several questions that can be used to evaluate the appropriateness of an assessment rubric relative to its intended purpose.

Content-Related Validity

1. Do the evaluation criteria address any extraneous content?
2. Do the evaluation criteria of the scoring rubric address all aspects of the intended content?
3. Is there any content addressed in the task that should be evaluated through the rubric, but is not?

Construct-Related Validity

1. Are all of the important facets of the intended construct evaluated through the scoring criteria?
2. Are any of the evaluation criteria irrelevant to the construct of interest?

Criterion-Related Validity

1. How do the scoring criteria reflect competencies that would suggest success on future related performances?
2. What are the important components of the future or related performance that may be evaluated through the use of the assessment instrument?
3. How do the scoring criteria measure the important components of the future or related performance?
4. Are there any facets of the future or related performance not reflected in the scoring criteria?

Reliability of assessment rubrics

The type of reliability most applicable to the use of scoring rubrics in the clinical setting is inter-rater reliability. It might also apply when multiple raters of classroom or written assignments are used. Inter-rater reliability, as stated earlier, refers to consistency in ratings given by two or more raters who are using the same scoring rubric. Statistical procedures for analysis of inter-rater agreement on rubrics using categorical scales include Cohen's kappa, for two independent raters, and *Fleiss's kappa,* for more than two raters (Garson, 2010). For interval-level scales, the ICC can be used.

Moskal and Leydens (2000, p. 7) emphasize the need to clearly state scoring rubrics so as to improve inter-rater reliability. They suggest the following criteria can be applied to evaluate the clarity of a rubric:

1. Scoring categories are well defined.
2. Differences between score categories are clear.
3. Two independent raters arrive at the same score for a given performance criteria based on the scoring rubric.

According to these authors, if any of these criteria are not met, then the unclear categories should be revised.

RUBRICS AND THE TIMELINESS OF FEEDBACK

An issue that can prove frustrating for both instructors and learners is the timeliness of feedback. In a study of 104 learners, Rucker and Thomson (2003) concluded that feedback, when given shortly after task completion, facilitated learning and helped learners make positive changes in subsequent work. While this finding may be intuitive, providing timely feedback can be a challenge, both in the practice arena and in the classroom. Well-developed rubrics can be an effective part of the teaching–learning process, as they inform learners about expectations and help instructors to provide timely, meaningful feedback that learners can utilize to improve their performance.

SUMMARY

Establishing educational effectiveness involves gathering credible evidence of learning and judging the extent to which the evidence is consistent with pre-established standards or learning objectives. Through clearly stated learning objectives, instructors inform learners what they expect them to know and do at the completion of a course of study. Rubrics and scoring guidelines are used to articulate the specific criteria that instructors use to judge how well learners meet learning expectations.

Assessment of learning occurs at several levels: learner, course, program, and institution. It can be competency based, formative or summative, direct or indirect, and qualitative or quantitative. Analysis of assessment data helps instructors shape curricular enhancements and improvements, provides learners with feedback about their own performance, guides instructors' decisions about learner progress, and provides evidence of quality to accrediting bodies and to those who have a stake in the educational program.

Quality assessments incorporate the properties of validity and reliability. Validity ensures accuracy of the assessment, whereas reliability ensures consistency across multiple assessment periods and multiple raters. Assessment measures that are valid, reliable, and linked with learning objectives will yield credible assessment data that instructors can use with confidence to provide evidence of learning.

REFERENCES

1. American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). Standards for teacher competence in educational assessment of students. Retrieved from http://www.unl.edu/buros/bimm/html/article3.html
2. American Nurses Association. (2010). Nursing: scope and standards of practice (2nd ed.). Retrieved from http://www.nursingworld.org/DocumentVault/NursingPractice/Draft-Nursing-Scope-Standards-2nd-Ed.aspx
3. Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., & Wittrock, M. C. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Addison Wesley Longman.
4. Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York: Susan Fauer.
5. Burns, M., Fager, J., Gumm, A., Haley, A., Krider, D., Linrud, J., & Webster, D. (n.d.). *Writing student learning outcomes for CMU programs*. Mount Pleasant, MI: Central Michigan University. Retrieved from https://academicaffairs.cmich.edu/caa/assessment/resources/toolkit/Determinestudentlearningoutcomes.doc
6. Chappuis, J. (2005). Helping students understand assessment. *Educational Leadership, 63*, 39–43.
7. Chirema, K. D. (2007). The use of reflective journals in the promotion of reflection and learning in post-registration nursing students. *Nurse Education Today, 27*, 192–202.
8. Garson, D. (2010). Reliability analysis. Retrieved from http://faculty.chass.ncsu.edu/garson/PA765/reliab.htm#rater
9. Gronlund, N., & Linn, R. (1990). *Measurement and evaluation in teaching and learning* (6th ed.). New York: Macmillan.
10. Holmboe, E. S., Juot, S., Chung, J., Norcini, J., & Hawkins, R. E. (2003). Construct validity of the MiniClinical evaluation exercise (MiniCEX). *Academic Medicine, 78,* 826–830.
11. Lacity, M., & Jansen, M. A. (1994). Understanding qualitative data: A framework of text analysis methods. *Journal of Management Information Systems, 11,* 137–160.
12. Leung, S. F., Mok, E., & Wong, D. (2008). The impact of assessment methods on the learning of nursing students. *Nurse Education Today, 28,* 711–719.
13. Mager, R. F. (1984). *Preparing instructional objectives* (2nd ed.). Belmont, CA: David S. Lake.
14. McCarthy, B., & Murphy, S. (2008). Assessing undergraduate nursing students in clinical practice: Do preceptors use assessment strategies? *Nurse Education Today, 28,* 301–313.
15. McCready, T. (2007). Portfolios and the assessment of competence in nursing: A literature review. *International Journal of Nursing Studies, 44*, 143–151.
16. McKelvie, S. J. (1992). Does memory contaminate test–retest reliability? *Journal of General Psychology, 119,* 59–72.
17. Miller, R., & Leskes, A. (2005). *Levels of assessment: From the student to the institution.* Washington, DC: Association of American Colleges and Universities.
18. Morrison, S., Smith, P., & Britt, R. (1996). *Critical thinking and test item writing.* Livingston, TX: Century II Printing.
19. Moskal, B. (2003). Recommendations for developing classroom performance assessments and scoring rubrics. *Practical Assessment, Research & Evaluation*, *8*(14). Retrieved from http://PAREonline.net/getvn.asp?v=8&n=14
20. Moskal, B., & Leydens, J. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research and Evaluation, 7*(10). Retrieved from http://Pareonline.net/getvn.asp?v=7&n=10
21. National Post Secondary Education Cooperative. (2002). Defining and assessing learning: Exploring competency-based initiatives. Report of the National Postsecondary Education Cooperative Working Group on Competency-Based Initiatives in Postsecondary Education. Retrieved from http://nces.ed.gov/pubs2002/2002159.pdf
22. Rucker, M. L., & Thomson, S. (2003). Assessing student learning outcomes: An investigation of the relationship among feedback measures. *College Student Journal, 37,* 400–405.
23. Scott, T. (2003). Bloom's taxonomy applied to testing in computer science classes. *Journal of Computing Sciences in Colleges, 19,* 267–274.
24. Zheng, A. Y., Lawhorn, J. K., Lumley, T., & Freeman, S. (2008, January 25). Application of Bloom's taxonomy debunks the "MCAT myth." *Science, 319,* 414–415.