

Astronomy Education Review

Volume 1, Oct 2001 - Jan 2002

Issue 1

The Role of Assessment in the Development of the College Introductory Astronomy Course

A "How-to" Guide for Instructors

by **Gina Brissenden**

American Astronomical Society

Timothy F. Slater

University of Arizona

Robert D. Mathieu

University of Wisconsin-Madison

National Institute for Science Education (NISE) College Level-One Team

University of Wisconsin-Madison

Posted: 02/26/02

The Astronomy Education Review, Issue 1, Volume 1:1-24, 2002

© , Gina Brissenden. Copyright assigned to the Association of Universities for Research in Astronomy, Inc.

Abstract

What is assessment? Why do it? Why do it in a particular way? This document addresses these important questions and provides a practical "how-to" guide for doing assessment. Assessment drives student learning; it is thus imperative that instructors conduct assessment in a manner that is well aligned with the instructor's goals for the course. This requires (a) that course goals be formalized, and (b) that the instructor have knowledge of various classroom assessment techniques and the kinds of course goals to which each of these assessment techniques is best suited. We briefly present several Classroom Assessment Techniques (CATs) that can be used to help instructors evaluate the extent to which course goals are being achieved, to help guide students toward desired learning outcomes, and to improve student self-evaluation of understanding. In addition, we outline a practical, generalized model for course development with which we demonstrate how to do assessment. For an on-line, user-friendly guide and resource to classroom assessment in college science courses, the reader is invited to visit the Field- Tested Learning Assessment Guide (FLAG) developed by the National Institute for Science Education (<http://www.wcer.wisc.edu/nise/cl1/flag>).

1. INTRODUCTION: WHAT IS ASSESSMENT? WHY DO IT? WHY DO IT IN A PARTICULAR WAY?

1.1. What is Assessment?

To many, the word "assessment" simply means the process by which faculty assign students grades. Assessment can be much more than this. Used properly, assessment provides the mechanism for gathering essential data about what our students are learning and about the extent to which we are meeting our teaching goals. Assessment is also a means for guiding and motivating students to be actively involved in their own learning. Indeed, **assessment drives student learning**: What we assess, and how we assess it, communicates to students what we want them to learn and how deeply we want them to understand it.

The types of assessment commonly used in first-year science, math, engineering, and technology (SMET) courses—giving students multiple-choice tests, for example—are typically intended only to inform students about their grade, or ranking, *after* they have received instruction (as opposed to, for example, ConceptTests, which give students real-time feedback *during* lecture). Given that this is the type of assessment our students most frequently encounter, and that it will eventually lead to their final course grades, students learn to study the content in our courses in an expeditious way that allows them to succeed in passing many first-year SMET courses without necessarily developing deep understanding of concepts (Deming, 2002). In fact, our approach to assessment drives the depth of student learning whether we want it to or not. The consequences of relying upon our "tried and true" assessment methods are profound, in that tests that measure only low-level cognitive skills may actively, even if unintentionally, promote superficial learning.

1.2. Why Do It?

Because of the role of assessment in driving student learning, choices about assessment (what to assess and how) should be made carefully and deliberately. Of course, we already "do assessment" to one extent or another, if only to decide what grades to assign. But too often we passively make "default" decisions regarding assessment without closely considering its connection to what we want our students to learn. On the one hand, this can lead to a dichotomy in our courses, where we say we want our students to learn one thing, but implicitly direct them to learn another. At the same time, this can lead to a disconnect between what our assessments are telling us about student learning, and what we infer from those assessments. (Example: Suppose we want our students to learn about the scientific process from our course, and that a student earns an 'A' on our end-of-term multiple-choice test. Does that 'A' student understand the scientific process?)

This dichotomy/disconnect can occur when we aren't clear to ourselves about what we want our students to learn, that is, when we haven't set clear learning goals. Assessment doesn't work as an isolated process, and isn't something to be done for its own sake. Assessment provides answers, but only if the questions are defined and explicitly stated. For us, the relevant questions that assessment can help answer will be of the sort: "Are my students learning what I want them to learn from this lecture?" and "Did my students learn what I wanted them to learn this semester?" Our students have equally important questions: "What am I expected to learn this week?" and "Did I learn what I was expected to learn this week?" Thus, for both instructors and students, assessment is only as effective as our goals are clear.

The importance of course goals--having them, articulating them, writing them down, and sharing them with students--cannot be overstated. For every course we teach, we must make decisions about what we want our students to know and be able to do by the end of the term (for examples of the most common learning goals among astronomy faculty, see Slater et al., 2001). Though we might not always consciously decide upon our goals, let alone formalize them in writing, we still make decisions about the assessment techniques we will employ (e.g., multiple-choice tests, essays, term papers, observing logs). The decisions we make about assessment direct students toward what they should learn.

Thus, simply "doing assessment" is not enough. If we wish to actively guide what our students learn, and how deeply they learn it, we must clearly decide what we want our students to take away from the course (set course goals), and then carefully choose our classroom assessment techniques accordingly (Anderson & Sosniak, 1994; National Research Council, 1996; Tobias & Raphael, 1997; Wiggins, 1998). Even if we aren't using assessment to answer the "What are my students learning?" questions, our students will nonetheless use our assessment choices to answer their "What am I supposed to learn?" questions, because their grades are on the line! By formalizing course goals, and then choosing appropriate assessment techniques, we can use the natural "leverage" that assessment provides to help guide our students toward what we actually want them to learn. Assessment drives student learning, and goals drive assessment. Measuring the attainment of course goals, and communicating those goals to our students, is why we do assessment.

1.3. Why Do It in a Particular Way?

Given the tight interrelationship between goals and assessment, to be most effective as instructors we should use classroom assessment techniques that are most appropriately suited to *actually measuring* attainment of our particular goals, or are *aligned with* our goals. The most commonly employed assessment method in first-year SMET courses is the multiple-choice test, typically administered at the end of a unit and/or at the conclusion of the entire course. While instructors often have very good operational reasons (e.g., time constraints) for using this assessment method, it may not be the best choice for actually measuring whether students are learning what they are expected to learn. Such tests are usually most effective at measuring students' fact-based knowledge and their ability to perform algorithmic problem-solving tasks. If our stated goals are that students be able to recite certain facts and solve simple algorithmic problems, then, in fact, the multiple-choice assessment technique is well aligned with the stated goals. However, if our goals include different student outcomes than these (e.g., an understanding of the scientific "process," a lifelong interest in the subject, the ability to critically analyze science in popular media, etc.), then this assessment technique will generally not provide us useful feedback about student attainment of these goals. Nor do these tests communicate to our students that this is what we expect them to learn, or provide useful feedback to students while there is still time for this to positively impact their learning.

As an alternative to multiple-choice tests, several other Classroom Assessment Techniques (CATs) that have been developed and field-tested have been found to be effective at both measuring student mastery of content and at giving students accurate cues about what they are expected to learn. Where time constraints are a concern, suitable CATs may be selected that are comparable to multiple-choice tests in terms of instructor effort. Later, we will present some of these techniques and discuss how instructors can select CATs that are best aligned with particular course goals.

However, before discussing specific CATs and their merits, we would like to take a step back and consider assessment from a broader perspective. We have already touched upon the connection between assessment and course goals, and assessment's role in driving student learning. But assuming you have set some course goals, how might you actually go about planning your assessment strategy? To answer this question, we need to look at assessment within the broader context of how a course is developed. Considerations about assessment are important to all aspects of the course development process, from formulating learning goals, to making decisions about course content and instructional methods, to measuring attainment of course goals. By examining this process and considering assessment's role in it, we can elucidate how assessment is done, and at the same time more fully explain why assessment should be done in the first place.

In what follows we consider a generalized model for course development, which we use to demonstrate "how to do assessment" using a variety of CATs. Not surprisingly, the model uses goals to determine the content, instructional methods, and CATs that are best suited for the course. We will see that assessment serves as the "feedback loop" wherein we evaluate the extent to which our choices about content and instructional methods are leading to the attainment of course goals. This will allow us to modify the content and instructional methods based on this evaluation. Because our focus here is on assessment specifically, this model will be presented in an idealized form, and certain details will only be sketched out. Subsequent papers in this series will flesh out this model further by examining various types of instructional methods (including "collaborative learning" and the use of instructional technologies); we invite readers to follow this series and to use this course development model as a guide and template for developing their own courses.

2. ASSESSMENT AS PART OF A GENERALIZED MODEL FOR COURSE DEVELOPMENT: A "HOW-TO" GUIDE FOR ASSESSMENT

2.1. Content, Instructional Methods, and Assessment

The three primary components of any course are the content, the instructional methods used to deliver the content, and the classroom assessment techniques (CATs) with which we evaluate whether students are achieving our learning goals. These three components are bound together by the overarching goals we set for the course. The course development model outlined here requires that course goals be formalized at the outset, which is to say that goals be clearly articulated. Ultimately, it is achievement of our goals by our students that is the standard against which the success of the course must be measured. In this context, the role of assessment is to measure the efficacy of our content and of our instructional methods *with respect to* student achievement of our goals. This is how content, instructional methods, and assessment are linked in this course development model.

While formalizing goals is an essential part of course development, it is only the first step. The path through the course development process can be envisioned as a "road map," with goals at the beginning, pointing the way, and with assessment telling us if we have reached our destination or if we need to retrace our steps. This course development "road map" (Figure 1) provides a detailed set of directions, with specific actions to be taken at several signposts along the way. Starting from formalizing course goals, the "directions" are as follows:

- *translate* goals into Measurable Student Outcomes
- *determine* Desired Levels of Expertise required to achieve outcomes
- *select* both Content and Classroom Assessment Techniques
- *choose and implement* Instructional Methods
- *conduct* Assessment and *evaluate*--were Measurable Student Outcomes realized?

Let's consider these steps in turn, with the goal of developing a fuller understanding of how to do assessment as part of the course development process.

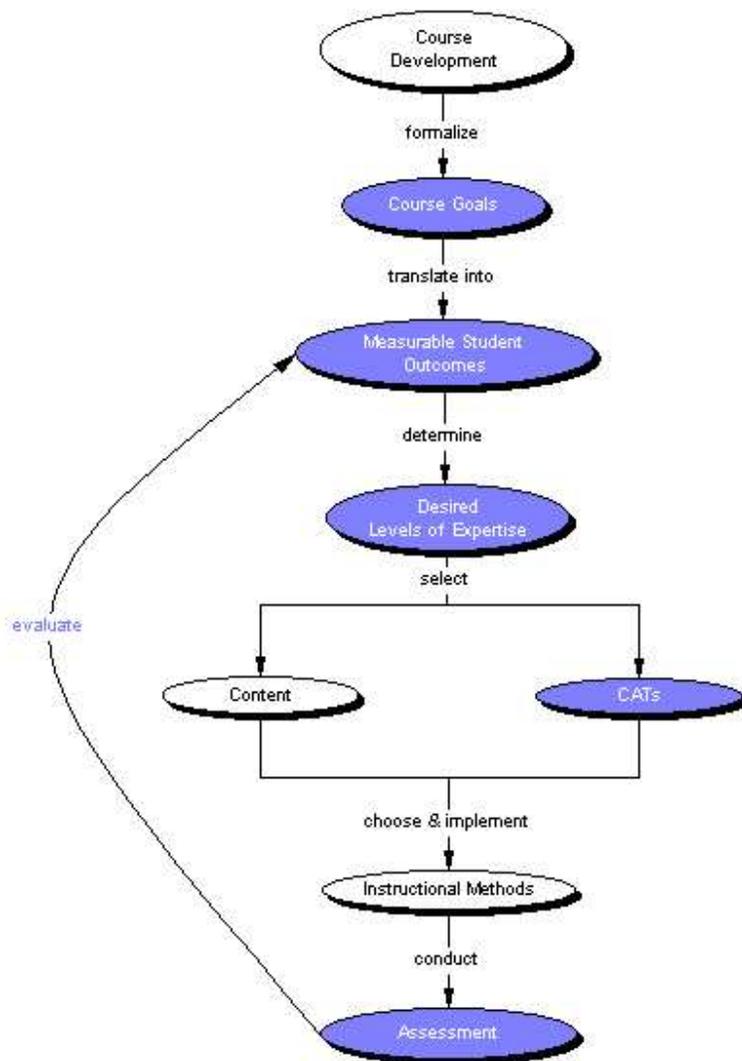


Figure 1. Roadmap of Course Development: A generalized model for course development. Steps related to doing assessment are highlighted.

2.2. Translating Course Goals into Measurable Student Outcomes

Assessment can measure the extent to which course goals have been achieved, but only if those goals are measurable. For the most part, however, course goals are too broad or too abstract to measure directly. This is one of the first difficulties often encountered with assessment in the course development process. For example, one course goal in an introductory astronomy course might be that "students understand the seasons." But how does one measure "understand"? This goal can be made more measurable by identifying specific learning outcomes one would expect from a student who "understands" the seasons. For example: The student can "define seasons" and can "distinguish the importance of different factors such as tilt and distance."

Thus, once goals have been formalized, the next step is to translate the often abstract language of course goals into a set of concrete *measurable student outcomes*. Measurable student outcomes are specific, demonstrable characteristics--knowledge, skills, values, attitudes, interests--that will allow us to evaluate, through assessment, the extent to which course goals have been met. For each course goal, identify the principal outcomes one would expect from a student who has achieved that goal, keeping in mind that our ability to measure the student achievement of course goals with CATs will be determined entirely on the basis of these measurable student outcomes. Figure 2 gives an example of translating a specific course goal (in the context of dental health) into measurable student outcomes. Of course, knowing what kinds of outcomes are actually measurable requires knowledge of the kinds of CATs that are available, and what each technique can and cannot measure. We discuss different CATs, and how to choose between them, below.

Dental Health 101

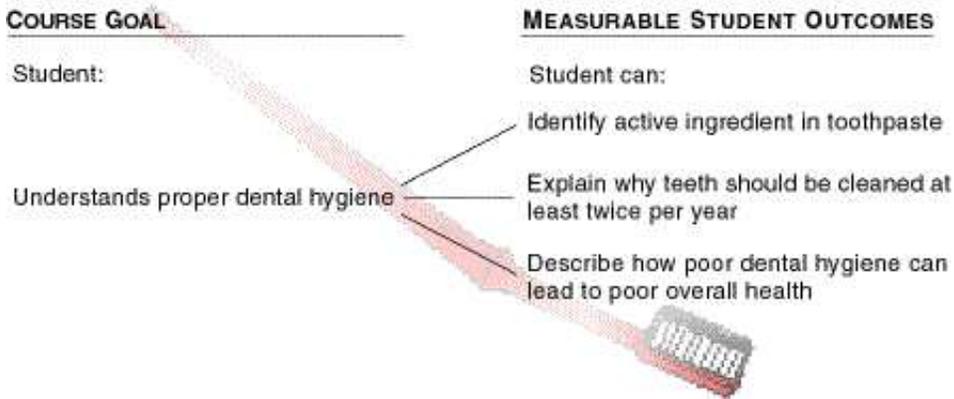


Figure 2. An example of translating a course goal into measurable student outcomes.

2.3. Determining Desired Levels of Expertise Required to Achieve Measurable Student Outcomes

Having translated course goals into measurable student outcomes, we are one step closer to selecting the CATs that will allow us to evaluate whether students are learning what we want them to learn. In order to select the CATs that are best suited for the course goals we have identified, it is advantageous to determine the levels of expertise that are required for achieving the measurable student outcomes that go with each course goal. The levels of expertise that we assign to measurable student outcomes are important because they are the factors that most directly determine the appropriate choices of CATs (as well as content and instructional methods) for the course.

What do we mean by "levels of expertise"? The various student outcomes that we assign to each course goal require different levels of mastery of course content. Some student outcomes require no more than students simply memorizing certain facts. However, many student outcomes require more sophisticated levels of understanding, or levels of expertise. Consider again the dental hygiene example above (Figure 2): The measurable student outcome of "knows the active ingredient in toothpaste" requires only that students *memorize* the correct answer (fluoride), while the outcome of "can describe how poor dental hygiene can lead to poor overall health" requires a much more sophisticated level of understanding, involving synthesis of multiple facts and concepts. Because *measurable student outcomes vary in the levels of expertise required to achieve them*, our CATs should be capable of assessing a variety of levels of expertise. In general, this means using a variety of CATs. Let's consider how to go about determining levels of expertise for our measurable student outcomes.

2.3.1. Bloom's Taxonomy of Educational Objectives

One of the most widely used ways of organizing levels of expertise is according to Bloom's Taxonomy of Educational Objectives (Bloom et al., 1994; Gronlund, 1991; Krathwohl et al., 1956). Bloom used a multi-tiered scale (Tables 1-4) to express the levels of expertise required to achieve different measurable student outcomes. Organizing measurable student outcomes in this way will allow us to select appropriate CATs for the course.

There are three Taxonomies. Which of the three to use for a given measurable student outcome depends upon the original goal to which the measurable student outcome is connected. There are *knowledge-based* goals, *skills-based* goals, and *affective* goals (affective means *values*, *attitudes*, and *interests*); accordingly, there is a separate taxonomy for each. Within each taxonomy, levels of expertise are listed in order of increasing complexity. Not surprisingly, measurable student outcomes that require the higher levels of expertise often require more sophisticated CATs.

The course goal in Figure 2--"student understands proper dental hygiene"--is an example of a *knowledge-based* goal. It is *knowledge-based* because it requires that the student learn certain facts and concepts. An example of a *skills-based* goal for this course might be "student flosses teeth properly." This is a *skills-based* goal because it requires that the student learn *how to do* something. Finally, an *affective* goal for this course might be "student cares about proper oral hygiene." This is an affective goal because it requires that the student's values, attitudes, or interests be affected by the course. Tables 1-4 introduce each of these taxonomies. Tables 1 and 2 are both examples of knowledge-based goals. Table 1 is based on knowledge about the functioning of a clock. We start with this example so that, regardless of specific content area, all instructors can share a common understanding of the levels of expertise. Tables 2-4 give astronomy-specific examples for each of the three taxonomies. The first three columns in each table are self-explanatory; the fourth column in Tables 2-4 will be explained in the section *Selecting Classroom Assessment Techniques*.

Table 1: Bloom's Taxonomy of Educational Objectives for Knowledge-Based Goals (General Example: Understanding How Clocks Work)

Level of Expertise	Description of Level	Example of Measurable Student Outcome
Knowledge	Recall, or recognition of terms, ideas, procedure, theories, etc.	Student can name the components of a simple clock.
Comprehension	Translate, interpret, extrapolate, but not see full implications or transfer to other situations, closer to literal translation.	Student knows the purpose of each component of a simple clock.
Application	Apply abstractions, general principles, or methods to specific concrete situations.	Student can describe how changing gear sizes will affect the precision of the clock.
Analysis	Separation of a complex idea into its constituent parts and an understanding of organization and relationship between the parts. Includes realizing the distinction between hypothesis and fact as well as between relevant and extraneous variables.	Given a malfunctioning clock, the student can design, and justify, a series of experiments to determine the cause of the malfunction.
Synthesis	Creative, mental construction of ideas and concepts from multiple sources to form complex ideas into a new, integrated, and meaningful pattern subject to given constraints.	Given a collection of clock parts, the student can design a clock that meets given specifications.
Evaluation	To make judgment of ideas or methods using external evidence or self-selected criteria substantiated by observations or informed rationalizations.	Given several novel alarm clock designs, the student can articulate the advantages and disadvantages of each.

Table 2: Bloom's Taxonomy of Educational Objectives for Knowledge-Based Goals (Astronomy Example: Understanding the Seasons)

Level of Expertise	Description of Level	Example of Measurable Student Outcome	CATs*
Knowledge	Recall, or recognition of terms, ideas, procedure, theories, etc.	When is the first day of Spring?	MCT, SALG
Comprehension	Translate, interpret, extrapolate, but not see full implications or transfer to other situations, closer to literal translation.	What does the summer solstice represent?	CT, CM, MTT, MCT, SALG, WR
Application	Apply abstractions, general principles, or methods to specific concrete situations.	Why are seasons reversed in the southern hemisphere?	CT, CM, CDT, MTT, MCT, SALG, Perf, WR
Analysis	Separation of a complex idea into its constituent parts and an understanding of organization and relationship between the parts. Includes realizing the distinction between hypothesis and fact as well as between relevant and extraneous variables.	What would Earth's seasons be like if its orbit were perfectly circular?	CT, CDT, IDI, MTT, Perf, Port, SR
Synthesis	Creative, mental construction of ideas and concepts from multiple sources to form complex ideas into a new, integrated, and meaningful pattern subject to given constraints.	Given a description of a planet's seasons, what would you propose its orbital and tilt characteristics to be?	IDI, Perf, Port, SR
Evaluation	To make judgment of ideas or methods using external evidence or self-selected criteria substantiated by observations or informed rationalizations.	What would be the important, and irrelevant, variables for predicting seasons on a newly discovered planet?	Port, SR

* Key: CT (ConceptTests), CM (Concept Maps), CDT (Conceptual Diagnostic Tests), IDI (In-Depth Structured Interviews), MTT (Mathematical Thinking Tasks), MCT (Multiple- Choice Tests), Perf (Performance Assessments), Port (Portfolio Assessments), SR (Scoring Rubrics), SALG (Student Self-Assessment of Learning Gains), WR (Weekly Reports)

Table 3: Bloom's Taxonomy of Educational Objectives for Skills-Based Goals (Astronomy Example: Ability to Use a Telescope)

Level of Expertise	Description of Level	Example of Measurable Student Outcome	CATs[†]
Perception	Uses sensory cues to guide actions.	Student realizes a "fuzzy" object in the night sky might be interesting to explore further.	WR
Set	Demonstrates a readiness to take action to perform the task or objective.	Student states that a telescope would be the most appropriate tool for investigating a "fuzzy" object in the night sky.	WR
Guided Response	Knows steps required to complete the task or objective.	Student can describe the steps involved in setting-up and aligning a telescope, and using it to find objects in the sky.	CM, IDI
Mechanism	Performs task or objective in a somewhat confident, proficient, and habitual manner.	Student can eventually locate three given galaxies.	Perf, SR
Complex Overt Response	Performs task or objective in a confident, proficient, and habitual manner.	Student can easily, and accurately, locate three given galaxies.	Perf, SR
Adaptation	Performs task or objective as above, but can also modify actions to account for new or problematic situations.	Student can easily, and accurately, select and locate three different galaxies on a partially cloudy night.	Perf, SR
Organization	Creates new tasks or objectives incorporating learned ones.	Student can successfully design, and host, a star party using a telescope. Student can make modifications to a telescope to allow for mounting of a heavy CCD camera.	Perf, Port, SR

[†] Key: CM (Concept Maps), IDI (In-Depth Structured Interviews), Perf (Performance Assessments), Port (Portfolio Assessments), SR (Scoring Rubrics), WR (Weekly Reports)

Table 4: Bloom's Taxonomy of Educational Objectives for Affective Goals (Astronomy Example: An Appreciation for Astronomy)

Level of Expertise	Description of Level	Example of Measurable Student Outcome	CATs[‡]
Receiving	Demonstrates a willingness to participate in the activity.	When I'm in class I am attentive to the instructor, take notes, etc. I do not read the newspaper instead.	AS
Responding	Shows interest in the objects, phenomena, or activity by seeking it out or pursuing it for pleasure.	I choose to allocate more free-time to watching astronomy programming on the Discovery Channel™.	AS
Valuing	Internalizes an appreciation for (values) the objectives, phenomena, or activity.	I believe it is important that the local high school support an astronomy club.	AS
Organizing	Begins to compare different values, and resolves conflicts between them to form an internally consistent system of values.	During vacations or business travels, I consistently include side trips to local planetaria or astronomy exhibits.	AS
Characterizing by a Value or Value Complex	Adopts a long-term value system that is "pervasive, consistent, and predictable."	I have joined, recruit for, and regularly attend functions of a local amateur astronomy club.	AS

[‡] Key: AS (Attitude Surveys)

To determine the level of expertise required for each measurable student outcome, you must first decide which of these three broad categories (knowledge-based, skills-based, or affective) the corresponding course goal belongs to. Then, using the appropriate Bloom's Taxonomy, look over the descriptions of the various levels of expertise. Determine which description most closely matches that measurable student outcome. As can be seen from the examples given in Tables 1-4, there are different ways of representing measurable student outcomes, e.g., as statements about students (Figure 2; Tables 1 & 3), as questions to be asked of students (Table 2), or as statements from the student's perspective (Table 4). You may find additional ways of representing measurable student outcomes; those listed in Figure 2 and in Tables 1-4 are just examples.

Bloom's Taxonomy is a convenient way to describe the degree to which we want our students to understand and use concepts, to demonstrate particular skills, and to have their values, attitudes, and interests affected. It is critical that we determine the levels of expertise that we are expecting our students to achieve because this will determine which CATs are most appropriate for measuring whether students are achieving the desired learning outcomes. Though the most common form of classroom assessment used in introductory college SMET courses--multiple-choice tests--might be quite adequate for assessing knowledge and comprehension (Levels 1 & 2; Tables 1 & 2), this type of assessment often falls short when we want to assess our students' knowledge at the higher levels of synthesis and evaluation (Levels 5 & 6) (Bloom et al., 1994; Tobias & Raphael, 1997). Multiple-choice tests also rarely provide information about achievement of skills-based goals. Similarly, traditional course evaluations, a technique commonly used to approximate affective assessment, do not generally provide useful information about changes in student values, attitudes, and interests.

Thus, commonly used assessment techniques, while perhaps providing a means for assigning grades, often do not provide us (nor our students) with useful feedback for determining whether students are attaining our course goals. Usually, this is due to a combination of not having formalized goals to begin with, not having translated those goals into outcomes that are measurable, and not using assessment techniques capable of measuring expected student outcomes given the levels of expertise required to achieve them. Using the model of course development presented here, we can ensure that our CATs are properly aligned with course goals--promoting intended learning.

Note that Bloom's Taxonomy need not be applied exclusively *after* course goals have been defined. Indeed, Bloom's Taxonomy and the descriptions associated with its different categories can help in the goals-defining process itself. In particular, Bloom's Taxonomy can be useful for ensuring that each of your learning goals includes (by way of measurable student outcomes) an appropriate range of levels of expertise (as in Figure 2). For example, learning goals that require relatively high levels of expertise should also have associated with them measurable student outcomes that require lower levels of expertise, providing students with "scaffolds" for building up to the higher-level aspects of the learning goal.

2.4. Selecting Course Content

At this point in the "road map," concrete decisions must be made about what will be included in the course content. As the instructor of the course, choices about content are entirely yours to make based upon what you want your students to take from the course. It would be beyond the scope of this document to attempt to discuss specific choices about content in detail. We do comment, however, that, here again, goals are paramount. If you are working from an existing syllabus (either your own or someone else's), take this opportunity to critically re-examine each component of the content *with respect to* your course goals. Are there topics in the syllabus that are not related to one or more course goals? Don't include content merely because "it's always been done that way" or because "it's important." If a topic is important, it will be reflected in your goals. If an "important" topic is not reflected in your course goals, you may wish to re-visit the goals themselves. In any case, each and every aspect of the content should connect clearly to course goals.

As a starting point for thinking about content in introductory science courses, the Society for College Science Teachers (see "Position Statement on Introductory College-Level Science Courses," available at <http://science.clayton.edu/scst/Courses.PDF>) has articulated the following general precepts:

An exemplary introductory science course should ... feature a carefully articulated sequence of topics that overtly illustrates, in a context of scientific inquiry, connections between concepts and principles germane to a course of study. The content and processes should not be all inclusive, rather they should represent the essential scientific information and skills of which students should become aware to function as scientifically literate and critically thinking adults. Accordingly, courses should emphasize the methodologies and logic used by scientifically literate people to investigate the world. Interdisciplinary connections between issues and principles of science, technology and society should be made where appropriate.

2.5. Selecting Classroom Assessment Techniques (CATs)

We are now at the point in the course development "road map" where we are ready to discuss the selection of specific CATs. Having gone through the previous steps of the course development model, decisions about which CATs to use can be made in a more informed manner, based upon specific measurable student outcomes and their associated levels of expertise. The CATs selected in this step will provide the feedback you need to evaluate the extent to which your course goals have been achieved. It is imperative that the CATs you select be properly matched with your measurable student outcomes.

To assist instructors in more readily incorporating new assessment techniques into their classrooms, the College Level-One (CL1) Team of the National Institute for Science Education (NISE) researched the most commonly used alternative assessment techniques and created an extensive Web site--the Field-Tested Learning Assessment Guide (FLAG)--to present eleven of these techniques. The CATs represented in the FLAG site have been tested in the field and are authored by national experts in the use of each particular technique. To be sure, the eleven CATs provided in the FLAG site are but a subset of innovative CATs available from a variety of resources (Adams & Slater, 2002; Angelo & Cross, 1993; Green, in press 2002; Siebert & McIntosh, 2001; Tobias & Raphael v1 & v2, 1997). They will, however, provide a good starting point for improving and implementing new assessment techniques into your courses.

3. CLASSROOM ASSESSMENT TECHNIQUES (CATS)

3.1. CAT Descriptions

Here we provide a brief description of each of the CATs represented in the FLAG site, but we encourage the reader to look to the FLAG site for a greater discussion of these techniques (<http://www.wcer.wisc.edu/nise/cl1/flag/>). Table 5 gives illustrative examples of the kinds of questions students might encounter when using each of these techniques.

Attitude Surveys (AS). Attitude surveys provide valuable information on student perceptions of their classroom experience. This includes general attitudes toward the course, the discipline, and their own learning. The results from this survey can also help you identify elements in your course which best support student learning. While attitudinal surveys may take many forms and address a range of issues, they typically consist of a series of statements with which students are asked to express their degree of agreement or disagreement, using a numerical scale.

ConceptTests (CT). With ConceptTests, the instructor obtains immediate feedback (during class) on the level of student understanding of a particular concept. Students obtain immediate practice in using SMET terminology and concepts. Students have an opportunity to enhance teamwork and communication skills. Many instructors have reported substantial improvements in class attendance and attitude toward the course. The instructor presents one or more questions during class involving key concepts, along with several possible answers. Students in the class indicate by, for example, a show of hands, which answer they think is correct. If most of the class has not identified the correct answer, students are given a short time in lecture to try to persuade their neighbor(s) that their answer is correct. The question is asked a second time by the instructor to gauge class mastery. Many variations on this general CAT exist.

Concept Maps (CM). Concept Maps assess how well students see the *big picture* surrounding a concept. They provide a useful and visually appealing way of illustrating students' conceptual knowledge. A Concept Map is a diagram of nodes, each containing concept labels, which are linked together with directional lines, also labeled. The concept nodes are sometimes arranged in hierarchical levels that move from general to specific concepts.

Conceptual Diagnostics Tests (CDT). Conceptual Diagnostics Tests are used to assess how well students understand key concepts in a SMET field prior to, during, and after instruction. These tests use items in a multiple-choice or short-answer format that are designed specifically to elicit common misconceptions.

In-Depth "Structured" Interviews (IDI). Using a handful of carefully selected students, In-Depth "Structured" Interviews enable assessment of the level of understanding your students have developed with respect to a series of well-focused, conceptually-related scientific ideas. This form of assessment provides feedback that is especially useful to instructors who want to improve their teaching and the organization of their courses. A "structured" interview consists of a series of well-chosen questions (and often a set of tasks or problems), which are designed to elicit a portrait of a student's understanding about a scientific concept or set of related concepts. The interview may be videotaped or audiotaped for later analysis.

Mathematical Thinking Tasks (MTT). Few faculty have difficulty finding or developing tools that assess the algorithmic mathematical techniques which they teach in SMET courses; a challenge which faculty do face, however, is finding ways to promote and assess the development of mathematical thinking--notably helping students know what to do when faced with problems that are not identical to the technical exercises they've already encountered in their course. Mathematical Thinking Tasks are designed to aid in the development of this problem-solving skill.

Multiple-Choice Tests (MCT). In any field of science, there exists a vocabulary, history, and basic knowledge base that constitute the foundation of the discipline. One efficient way to measure students' abilities to recall and identify these basic constituents is the oft-used multiple-choice test. The most common multiple-choice test items are constructed with an incomplete sentence as the prompt, or stem, which is followed by several choices. One of these choices is a most appropriate completion to the stem, whereas the other three choices, called distracters, represent common mistakes that students make. Multiple choice items are quick and easy to grade, but often difficult to write well.

Performance Assessment (Perf). Although facts and concepts are fundamental in any undergraduate SMET course, knowledge of methods, procedures, and analysis skills that provide context are equally important. Student growth in these latter facets proves somewhat difficult to evaluate, particularly with conventional multiple-choice examinations. Performance assessments, used in concert with more

traditional forms of assessment, are designed to provide a more complete picture of student achievement. Performance assessments are designed to judge student abilities to *use* specific knowledge and research skills. Most performance assessments require students to manipulate equipment, to solve a problem, or to make an analysis. Rich performance assessments reveal a variety of problem-solving approaches, thus providing insight into a student's level of conceptual and procedural knowledge.

Portfolio Assessment (Port). Portfolio Assessment strategies provide a structure for long-duration, in-depth assignments. The use of portfolios transfers much of the responsibility of demonstrating mastery of concepts from the instructor to the student. Student portfolios are a collection of evidence, prepared by the student and evaluated by the instructor or teaching assistants, that demonstrate mastery, comprehension, application, and synthesis of a given set of concepts. To create a high-quality portfolio, students must organize, synthesize, and clearly describe their achievements, and effectively communicate what they have learned.

Scoring Rubrics (SR). Has a student ever said to you regarding an assignment, "But, I didn't know what you wanted!" or "Why did her paper get an 'A' and mine a 'C'?" Students must clearly understand the level of performance we expect them to achieve in course assignments, and importantly, the criteria we use to determine how well they have achieved those goals. A Scoring Rubric, though not technically itself a CAT (it is used in conjunction with a CAT), provides a readily accessible way of communicating our goals to students as well as communicating the criteria we use to discern how well students have reached them. Rubrics (or "scoring tools") are a way of describing evaluation criteria (or "grading standards") based on the expected outcomes and performance of students. Typically, rubrics are used in scoring or grading written assignments or oral presentations; however, they may be used to score any form of student performance. Each rubric consists of a set of scoring criteria and point values associated with these criteria. In most rubrics the criteria are grouped into categories so the instructor and the student can discriminate among the categories by level of performance. In classroom use, the rubric provides an "objective" external standard against which student performance may be compared.

Student Self-Assessment of Learning Gains (SALG). Strategies that allow for Student Self-Assessment of Learning Gains can spotlight those elements in the course that best support student learning and those that need improvement. This instrument is a powerful tool, can be easily individualized, provides instant statistical analysis of the results, and facilitates formative evaluation throughout a course. (Note: A subsequent article in this series will describe the SALG technique and its use in greater depth.)

Weekly Reports (WR). Weekly Reports provide rapid feedback about what students think they are learning and what conceptual difficulties they are experiencing. Weekly Reports are short papers written by students each week, in which they typically address three questions: "What did I learn this week?", "What questions remain unclear?", and "What questions would you ask your students, if you were the professor, to find out if students understood the material?"

The above capsule summaries are intended only to provide a cursory overview of the assessment resources available to you in the form of ready-to-use CATs at the FLAG Web site. The FLAG was designed with the on-line user in mind. We invite you to jump to the FLAG site to learn more about the CATs and, more importantly, how to use them. For each CAT you will find:

- **Focus Questions:** Overview of strategy, general requirements, and limitations of implementation.
- **Description:** Succinct but thorough introduction, including time estimates for implementation.
- **Purposes:** Indications of appropriate usage.
- **Limitations:** Contra-indications and potential problems.
- **Teaching Goals:** List of course goals addressed by the strategy.
- **Suggestions for Use:** Friendly "tips" from an experienced user.
- **Step-by-Step:** Explicit directions for implementation.
- **Variations:** Alternative uses and elaborations.
- **Analysis:** Making sense of the data; uses in evaluation.
- **Pros and Cons:** Advantages and disadvantages.
- **Theory and Research:** Conceptual and empirical foundations.
- **Links:** URLs or e-mail addresses of CAT authors for direct contact by users.
- **Sources:** Books, papers, related Web sites.
- **Author's Story:** Personalized description of author and how (s)he came to use strategy.

Table 5: Illustrative Examples of Classroom Assessment Techniques Featured on the FLAG Site

Classroom Assessment Technique	Illustrative Example	Time Estimates (Prep Time and Class Time)
Attitude Surveys	<p>Astronomy is contributing new knowledge that is important to society.</p> <p>(strongly agree <- 1 2 3 4 5 -> strongly disagree)</p>	<p>Prep: Very little time is needed to use a valid, existing survey. Large amounts of time are required to develop a survey that is reliable and measures what is intended.</p> <p>Class: Varies with length, but rarely more than 20 minutes.</p>
ConcepTests	<p>Answer first by yourself, then with a partner: Which of the following makes the determination of Hubble's Constant most difficult?</p> <p>a) unknown recessional velocities b) unknown galactic distances c) existence of dark matter d) lack of federal funding</p>	<p>Prep: Some time is needed to create ConcepTests. For some disciplines, hundreds of sample questions exist on Web sites as a time-saving resource.</p> <p>Class: ConcepTests typically last from less than a minute to several minutes.</p>

<p>Concept Maps</p>	<p>Create a concept map showing the evolution of matter in the Universe, from the products of the Big Bang to the iron in your blood.</p>	<p>Prep: Minimal if students construct maps; large for designing "fill-in" maps.</p> <p>Class: Varies depending on whether student-constructed or "fill-in," but rarely more than 20 minutes.</p>
<p>Conceptual Diagnostics Test</p>	<p>A sample item from the Astronomy Diagnostics Test (ADT): A flag pole in Denver will have no shadow at noon</p> <p>a) on the first day of spring b) on the first day of summer c) every day d) never</p>	<p>Prep: Minimal for using available tests; moderate for designing your own questions.</p> <p>Class: At least 30 minutes for a complete test.</p>
<p>In-Depth "Structured" Interviews</p>	<p>The instructor asks students to explain the meaning, and importance, of the astronomical "distance ladder."</p>	<p>Prep: Several hours required to develop a set of good questions, tasks, and problem sets. Additional time to locate appropriate props and recording equipment, if desired.</p> <p>Class: One-on-one or small group interviews may be conducted in less than an hour in your office or other convenient "private space." Some practice will reduce the time required to conduct a good interview.</p>
<p>Mathematical Thinking Tasks</p>	<p>A question related to scale: Suppose a chain is made from a million paper clips. How far will it stretch? Choose suitable units for your answer. Include your assumptions and reasoning.</p>	<p>Prep: Minimal if using available tasks (e.g., available on line).</p> <p>Class: Some tasks take 5 minutes, others as much as 45 minutes.</p>
<p>Multiple-Choice Tests</p>	<p>When our Sun depletes its available fuel supply, it will eventually become a</p> <p>a) supernova b) black hole c) white dwarf d) neutron star</p>	<p>Prep: Minimal if using existing questions (e.g., from a test-bank).</p> <p>Class: Depends on number of questions asked, but typically one class period a few times during the term.</p>

<p>Performance Assessment</p>	<p>Set up a telescope, show your instructor three objects, and describe interesting attributes of each.</p>	<p>Prep: Medium. Class: 10-40 minutes depending on complexity of task.</p>
<p>Portfolio Assessment</p>	<p>For your course "portfolio," select four assignments (homework, term papers, observation logs, exams) you completed this semester that most clearly demonstrate your knowledge of astronomy. Compose a two-page letter explaining why these materials clearly demonstrate mastery. Include the four assignments and this letter in the portfolio.</p>	<p>Prep: Minimal, after the course learning objectives have been clearly identified. Can be high if multiple graders are to be trained (e.g., graduate teaching assistants) when used in large classes. Class: None.</p>
<p>Scoring Rubrics</p>	<p>For an observing log, students are scored as follows:</p> <p>ADVANCED: All aspects of a PROFICIENT, plus makes reasonable hypotheses about future observations that can be tested.</p> <p>PROFICIENT: Student log clearly and correctly identifies object observed, uses a sketch that shows position of the object, and includes descriptive notes that relate this observation to other observations.</p> <p>NEARING PROFICIENT: Student log clearly and correctly identifies object but does not adequately describe its position or is missing components required for a PROFICIENT score.</p> <p>NOVICE: Student log insufficiently provides enough information to clearly demonstrate that student made the assigned observation correctly or at expected level of participation.</p>	<p>Prep: Variable. As students use rubrics, they become better writers and oral presenters; hence the time instructors spend evaluating students' work is reduced. Class: Variable. As students use rubrics, they become better writers and oral presenters; hence the time instructors spend evaluating students' work is reduced.</p>

<p style="text-align: center;">Student Self-Assessment of Learning Gains</p>	<p>Rate your understanding of each of the following using a scale of one (no understanding) to five (complete understanding): phases of Venus, seasons of Mars, and causes of supernovae. Further, rate the effectiveness of each of the following in helping you learn astronomy using a scale of one (did not help at all) to five (was essential in helping me learn): homework problems, pop quizzes, lectures, and reading assignments.</p>	<p>Prep: Time is needed to: clarify and prioritize class learning objectives and the related activities that the teacher wishes to be evaluated; check which existing questions express these and which need to be edited or added. No instructor time is needed to administer the survey, collect, and analyze the resultant data.</p> <p>Class: Instrument can be given in or out of class. It takes 10-15 minutes to complete the sample instrument.</p>
<p style="text-align: center;">Weekly Reports</p>	<p>Submit a one-page report that explains the two most important concepts covered in class this week, which topic you are finding the most difficult to understand, and what you think would make a good test question from this unit.</p>	<p>Prep: Minimal. Questions may be written on blackboard or provided in hard copy form.</p> <p>Class: None; done at home.</p>

Though many of these CATs could be adapted to assess measurable student outcomes at any of the Bloom Taxonomy levels of expertise, the last column of Tables 2-4 gives a listing of appropriate CATs to help you get started. In addition, for more in-depth, classroom-ready, astronomy-specific examples of the Attitude Surveys, Concept Maps, Conceptual Diagnostic Tests, and Student Self-Assessment of Learning Gains CATs, we invite the reader to visit the Tools section of the FLAG site.

(While the CATs represent broadly applicable assessment techniques, the tools are discipline-specific instruments that are "nested" within CATs (e.g., Conceptual Diagnostic Tests represents a general category of assessment technique; *The Astronomy Diagnostic Test* is a specific tool for use in an introductory astronomy course). Since some instructors may be looking for *convenient, time-saving, discipline-specific tools or instruments* that may be implemented directly with only a modicum of additional effort, we have devised a database of appropriate tools. The database can be sorted by discipline or CAT, or searched by discipline, purpose, or CAT.)

3.2. Choosing and Implementing Instructional Methods

Having our course goals formalized, having translated those goals into measurable student outcomes and assigning to each appropriate levels of expertise, and having selected the course content and CATs, we are in the best possible position to teach effectively. That is, we are at the point in the course development "road map" where we choose and implement the instructional methods that will best deliver the course content.

As with choosing CATs, the choice of instructional methods must be guided by our course goals and, perhaps even more so, by our expected levels of expertise associated with measurable student outcomes. For example, suppose that two of the course goals in an introductory engineering course are (1) that

students learn how to "design simple devices that satisfy realistic constraints" and (2) that students "can work effectively as part of a design team." A common measurable student outcome for these goals might be that "students, working in a team, can design a device, using simple raw materials, that protects an egg when dropped from a height of fifteen feet." This measurable student outcome is at the "Organization" level of expertise (Bloom's Taxonomy of Educational Objectives for Skills-Based Goals; Table 3) because it requires students to "create new tasks or objectives incorporating learned ones." Traditional lecturing alone would not be a sufficient instructional method in this case. Instead, an instructional method that emulates teamwork and that promotes creative thought would be more appropriate. That is, a more collaborative instructional method is called for.

A variety of instructional methods have been developed for guiding students to the different levels of expertise represented by the goals of our course. One commonly used instructional method--collaborative learning--is described, in detail, at the NISE Collaborative Learning Web site (<http://www.wcer.wisc.edu/nise/cl1/CL/>). In fact, you will find that collaborative learning instructional methods are appropriate and useful for a wide variety of goals, outcomes, and levels of expertise, and can be used in conjunction with many of the CATs presented above. A more detailed discussion of Collaborative Learning instructional methods, as well as a discussion of Learning Technologies, will be presented in future papers.

3.3. Conducting Assessment and Evaluating Attainment of Goals: Closing the Feedback Loop

It is in this final step of the course development process that we harness the power of the data provided by the CATs used during the implementation of the course. Some of these assessment data may be used for assigning grades to our students. Ultimately, however, the real value of these assessment data comes when we use them for improving the course and student learning. That is, our assessment data provide us with critical feedback for evaluating what we've done--what works and what doesn't. Depending upon the CATs we have chosen, this feedback may be used either at the end of the course (to summarize the efficacy of our course development efforts) or along the way (to inform our course development efforts in progress). When assessment is used to evaluate the course in summary fashion at the end, it is called *summative*. When assessment is used to modify the course *while it is in progress*, it is called *formative*. Either way, the point of assessment is to give us the information we need for evaluating attainment of our course goals.

How, specifically, do we perform this evaluation? By what criteria do we know if we have achieved our goals? Our measurable student outcomes are the key: If these outcomes are realized, we will know that we have attained our course goals. Look at the assessment data. Did your students achieve the hoped-for outcomes, and at the desired levels of expertise? Using the engineering example from above, perhaps the egg survived but you learn from student weekly reports that one student in the team did all of the work, i.e., the outcome related to teamwork was not realized. How might you modify the course to better foster effective teamwork? Perhaps students needed more guidance on how to work collaboratively; consider how the content and/or instructional methods might be changed to accomplish this. Or perhaps it is the teamwork *goal* itself that needs refining. It is this important, evaluative step that allows you to determine the extent to which you are reaching your course goals and to decide if there are changes you would like to make.

We close by noting that the terms *assessment* and *evaluation* are often incorrectly used synonymously. *Assessment* is the collecting of data to inform both the instructor and the student as to how the course is progressing (*formative assessment*) or how it has ended (*summative assessment*). Assessment involves gathering data via one or more CATs. *Evaluation* is what we do with these data once we have them. Once we have collected the assessment data, it is up to us to judge the efficacy, and value, of our instructional methods, the content of our course, and the achievement of our course goals.

4. SUMMARY

We have discussed the importance of assessment: what it is, why we should do it, and why we should do it in a particular way. Assessment is much more than the process by which we assign grades. Assessment is a means for providing critical feedback to both the instructor and her/his students. For us, assessment provides the data we need to evaluate the efficacy of our course (i.e., instructional methods and content) with respect to stated learning goals. For our students, our classroom assessment techniques (CATs) communicate--perhaps more loudly than words--what those learning goals are. Thus, we have seen that there is an important link between assessment and goals: Assessment drives student learning; our goals should drive our assessments. By setting goals, translating these goals into measurable student outcomes, and choosing appropriate CATs, we can promote the type of student learning we want.

Indeed, given the power of assessment for guiding students toward desired learning outcomes, one might say that assessment is the most seriously "underrated" tool at the instructor's disposal. We already do assessment of one sort or another (typically in the form of multiple-choice tests and student course evaluations), but are we using the most appropriate CATs given our course goals? The answer to this question is obviously a function of the individual instructor and her/his specific goals.

To facilitate faculty in choosing, and employing, appropriate CATs for their course goals, the National Institute for Science Education (NISE) College Level-One Team has created the Field-Tested Learning Assessment Guide (FLAG) Web site. The FLAG site offers broadly applicable, self-contained, modular CATs and discipline-specific tools for SMET instructors interested in alternative approaches to assessing student learning, skills, and attitudes. Each CAT has been developed, tested, and refined by recognized experts in real college and university classrooms. The FLAG site also contains a section to help you select the most appropriate CATs for your course goals and links to additional resources.

We have also outlined a general model for course development. The model begins with course goals, reflecting the role that goals play throughout the course development process, especially with regards to conducting assessment. Our goal in presenting this model has been twofold. First, we hoped to provide a useful template for course development that we believe readers can use for the development of introductory (and other) astronomy courses. Our focus here has been on assessment; subsequent papers in this journal will build upon this template by discussing a variety of instructional methods. Second, and more importantly, we sought to demonstrate "how to do assessment"; course development is perhaps the most natural context for seeing how assessment is done. While instructors can significantly enhance their courses by simply incorporating some of the CATs presented here (so long as they are aligned with course goals), these CATs are most useful during the initial development of a new course or as part of a systematic re-evaluation of an existing one.

Resources

Field-Tested Learning Assessment Guide (FLAG) - NISE Assessment Web site:

<http://www.wcer.wisc.edu/nise/cl1/flag/>

CATs: <http://www.wcer.wisc.edu/nise/cl1/flag/cat/cat.htm>

Tools: <http://www.wcer.wisc.edu/nise/cl1/flag/tools/tools.htm>

Goals: <http://www.wcer.wisc.edu/nise/cl1/flag/goals/goals.htm>

Resources: <http://www.wcer.wisc.edu/nise/cl1/flag/resource/resource.htm>

Collaborative Learning - NISE Collaborative Learning Web site: <http://www.wcer.wisc.edu/nise/cl1/CL/>

Learning Technologies - NISE Learning Technologies Web site: <http://www.wcer.wisc.edu/nise/cl1/ilt/>

References

Adams, J. P., & Slater, T. F. 2002, *Astronomy Teaching that Focuses on Learning*, Boston: Prentice Hall.

Angelo, T. A., & Cross, K. P. 1993, *Classroom Assessment Techniques: A Handbook for College Teachers*, San Francisco: Jossey-Bass.

Bloom, B. S., et al. 1994, Taxonomy of Educational Objectives, the Classification of Educational Goals, Handbook I: Cognitive Domain, in *Bloom's Taxonomy: A Forty-Year Retrospective*, L. W. Anderson & L. A. Sosniak (Eds.), Chicago: University of Chicago Press.

Deming, G. L. 2002, Results from the Astronomy Diagnostics Test, *Bulletin of the American Astronomical Society*, 33(4), 80.01,

Green, T. 2002, *ConceptTest for Introductory Astronomy*, Boston: Prentice Hall.

Gronlund, N. E. 1991, *How to Write and Use Instructional Objectives*, 4th Ed., New York: Macmillan Publishing Co.

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. 1956, *Taxonomy of Educational Objectives, the Classification of Educational Goals, Handbook II: Affective Domain*, New York: David McKay Co., Inc.

Linn, R. L. 1995, *Measurement and Assessment in Teaching*, 7th Ed., Englewood Cliffs, N.J.: Merrill.

National Research Council. 1996, *National Science Education Standards*, Washington, D.C.: National Academy Press.

Siebert, E. D., & McIntosh, W. J. (Eds.) 2001, *College Pathways to the Science Education Standards*, Arlington, Va.: NSTA Press.

Slater, T., Adams, J. P., Brissenden, G., & Duncan, D. 2001, What Topics are Taught in Introductory Astronomy Courses?, in *The Physics Teacher*, 39(1), 52.

Tobias, S., & Raphael, J. 1997, *The Hidden Curriculum-Faculty-Made Tests in Science, Part 1: Lower-Division Courses*, New York: Plenum Press.

Tobias, S., & Raphael, J. 1997, *The Hidden Curriculum-Faculty-Made Tests in Science, Part 2: Upper-Division Courses*, New York: Plenum Press.

Wiggins, G. P. 1998, *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*, San Francisco: Jossey-Bass.

ÆR

1 - 24