# Table of Contents

**Contributors**

Preface

## Section 1. General Assessment Topics

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 1:</td>
<td>Survey Instrument Validation: The First Commandment of Educational Research</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td><em>Cynthia Cudaback</em></td>
<td></td>
</tr>
<tr>
<td>Chapter 2:</td>
<td>Building a Culture of Faculty-Owned Assessment</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td><em>Don Haviland, Karol Dean, and Eleanor D. Siebert</em></td>
<td></td>
</tr>
<tr>
<td>Chapter 3:</td>
<td>Quantitative Assessment of Student Learning in Large Undergraduate Science Classrooms: Approaches and Caveats</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td><em>Anneke M. Metz</em></td>
<td></td>
</tr>
<tr>
<td>Chapter 4:</td>
<td>Means of Linking Research to Practice in Organizing a Course on Student Assessment</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td><em>HsingChi von Bergmann</em></td>
<td></td>
</tr>
</tbody>
</table>

## Section 2. Assessment in the College Classroom

### A. Traditional Assessments

| Chapter 5: | Writing/Using Multiple-Choice Questions to Assess Higher-Order Thinking | 35   |
|           | *Kerry L. Cheesman*                                                    |      |
| Chapter 6: | Tips on Classroom Assessment: How to Teach Our Students to Take Multiple-Choice Exams | 43   |
|           | *Linda Tichenor*                                                       |      |
| Chapter 7: | Better Multiple-Choice Assessments                                     | 45   |
|           | *William J. Straits and Susan Gomez-Zwiep*                             |      |
Table of Contents

B. Alternative Assessments

Chapter 8: Alternative Forms of Assessment for the College Science Laboratory
Kerry L. Cheesman

Chapter 9: Assessment of Students’ Learning Through the Creation of Scientific Posters
Kerry L. Cheesman and Nancy J. Swails

Chapter 10: Using Electronic Portfolios for Assessment in College Science Courses: Instructor Guidelines and Student Responses
Jerry A. Waldvogel

C. Teacher Education Alternative Assessments

Chapter 11: Wetscience: A Means of Assessing Math, Science, and Technology Incorporation Into a Service Learning Outreach Program
Mark A. Gallo

Chapter 12: Gauging the Nature of Science (NOS): An Alternate Form of Assessment
Thomas J. Melvin

Chapter 13: Authentic Assessment: Using 5-E Lesson Plan Development to Evaluate Science Content Learning With Preservice Teachers
Holly Travis
Section 3. How-To Section
Successful Classroom-Tested Practices and Instructions and Rubrics for Their Implementation

Chapter 14: Formative Assessment With Student Remotes and E-mail
Robert A. Cohen

Chapter 15: Peer Assessment: Value, Fears, Headaches, and Success
Anne Coleman

Chapter 16: Working With Student Engagement
Grace Eason

Chapter 17: Promoting Student Reflection on Exams
Grace Eason

Chapter 18: Hypothesis Modification Activity
Brian J. Rybarczyk and Kristen Walton

Chapter 19: Exam Corrections and Analysis, Student Perspective
Kathryn H. Sorensen

Chapter 20: Exam Analysis, Instructor Perspective
Kathryn H. Sorensen

Chapter 21: Inquiry-Based Labs: The Scientific Report
Bonnie S. Wood

Chapter 22: Student-Authored Book Reviews
Bonnie S. Wood

Chapter 23: Student-Led Teaching Models
Bonnie S. Wood
### Table of Contents

**Section 4. General Practices to Improve Assessment**

**Chapter 24:** Eleven Assessment Lessons Learned at the Gate  
*Mary H. Brown*  
123

**Chapter 25:** Developing Assessment Performance Indicators  
*Peter Kandlbinder*  
129

**Chapter 26:** Practices That Jeopardize Bona Fide Student Assessment  
*Thomas R. Lord*  
137

**Chapter 27:** Varied Assessment: A Brief Introduction  
*William J. Straits and R. Russell Wilke*  
147

**Chapter 28:** Assessments That Assist in Motivating Students  
*Ellen Yerger*  
151

**Index**  
155
Contributors

Brown, Mary H.
Lansing Community College
MC 5400—Science
411 N. Grand Ave.
Lansing, MI 48901
Brownm@lcc.edu

Cheesman, Kerry L.
Capital University
Columbus, OH
kcheesma@capital.edu

Coleman, Anne
Department of Life and Physical Sciences
Cabrini College
610 King of Prussia Rd.
Radnor, PA 19087
annecoleman@cabrini.edu

Cohen, Robert A.
Department of Physics
East Stroudsburg University
East Stroudsburg, Pennsylvania
rcohen@po-box.esu.edu

Cudaback, Cynthia
Marine Earth and Atmospheric Science
Campus Box 8208
North Carolina State University
cynthia_cudaback@ncsu.edu

Dean, Karol
Mount St. Mary’s College
12001 Chalon Rd.
Los Angeles, CA 90049
kdean@msmc.la.edu

Eason, Grace
University of Maine at Farmington
173 High St., Preble Hall
Farmington, ME 04938
gseason@maine.edu

Gallo, Mark A.
Niagara University
DePaul Hall,
Niagara University, NY 14109
mgallo@niagara.edu

Gomez-Zwiep, Susan
California State University, Long Beach
Department of Science Education
Long Beach, CA 90840
sgomezwp@csulb.edu

Haviland, Don
California State University, Long Beach
1250 Bellflower Blvd.
Long Beach, CA 90808-2201
dhavilan@csulb.edu

Kandlbinder, Peter
Institute for Interactive Media and Learning,
University of Technology, Sydney
Peter.Kandlbinder@uts.edu.au

Lord, Thomas R.
Department of Biology
Indiana University of Pennsylvania
Indiana, PA 15705
trlord@iup.edu

Melvin, Thomas J.
Biology Department
Indiana University of Pennsylvania
Indiana, Pennsylvania 15705
t.j.melvin@iup.edu

Metz, Anneke M.
Department of Cell Biology and Neuroscience
Montana State University
Bozeman, MT 59717
anneke@montana.edu
Contributors

Rybarczyk, Brian J.
The Graduate School
University of North Carolina/Chapel Hill
CB#4010
Chapel Hill, NC 27599-4010
brybar@unc.edu

Siebert, Eleanor D.
Mount St. Mary’s College
12001 Chalon Rd.
Los Angeles, CA 90049
esiebert@msmc.la.edu

Sorensen, Kathryn H.
Biology Department
American River College
4700 College Oak Dr.
Sacramento, CA 95841
SorensKH@arc.losrios.edu

Straits, William J.
California State University, Long Beach
Department of Science Education
Long Beach, CA 90840
wstraits@csulb.edu

Swails, Nancy J.
Capital University
Columbus, OH 43209
swails@capital.edu

Tichenor, Linda
Department of Biology
University of Arkansas
Fort Smith, Arkansas 72913
lticheno@uafortsmith.edu

Travis, Holly
Department of Biology
Indiana University of Pennsylvania
Indiana, Pennsylvania 15705
holly.travis@iup.edu

von Bergmann, HsingChi
EDT 726, Faculty of Education
University of Calgary
Calgary, AB, Canada T2N 1N4
hsingchi@ucalgary.ca

Waldvogel, Jerry A.
Department of Biological Sciences
Clemson University
Clemson, SC 29634-0314
waldvoj@clemson.edu

Walton, Kristen
Department of Biology
Missouri Western State University
4525 Downs Dr.
St. Joseph, MO 64507
kwalton@missouriwestern.edu

Wilke, R. Russell
Angelo State University
Department of Biology
ASU Station - 10890
San Angelo, TX 76909
russell.wilke@angelo.edu

Wood, Bonnie S.
University of Maine at Presque Isle
181 Main St.
Presque Isle, ME 04769
bonnie.s.wood@umpi.edu

Yerger, Ellen
Department of Biology
Indiana University of Pennsylvania
114 Weyandt Hall
Indiana, Pennsylvania 15701
ellen.yerger@iup.edu
Welcome to the sixth in the series of SCST monographs created by the members of the Executive Board of the Society for College Science Teachers in cooperation with the National Science Teachers Association. This document covers an extremely important and often controversial topic, that of evaluating the value of students and professorial works. The jointly sponsored monograph has been three years in the making with the initial agreement with the National Science Teachers Association taking place in the fall of 2006.

Each submission in this monograph was reviewed by at least two members on the Editorial Board with the published authors responding to reviewers critiques and providing the final proofing of their own entry. Articles were selected on the quality of the writing and their contribution to the value and importance of assessment in a college science setting.

The monograph examines assessment issues from several different viewpoints and is broken into several chapters. The first section deals with general assessment topics such as validation of survey instruments and creating a culture for faculty-owned assessment. The second section concerns traditional and alternative forms of assessment in both science and the science education classroom. The third section presents a series of how-to assessment practices that have been successfully utilized in the field. Finally, the fourth section provides a series of tips to enhance assessment in the college science classroom.

The editors would like to thank all the contributors to the monograph. The quality of the initiative is indicative of the time and energy they put into this work.

Thomas R. Lord
Indiana University of Pennsylvania
trlord@iup.edu

Donald P. French
Oklahoma State University
dfrench@okstate.edu

Linda W. Crow
Montgomery College
linda.w.crow@nhmccd.edu

Acknowledgments
The editors wish to thank Holly Travis for her tireless effort and dedication to the construction of this document. Thanks also to Ellen Yerger and Tom Melvin for their help in making this monograph a success.
Chapter 5

Writing/Using Multiple-Choice Questions to Assess Higher-Order Thinking

Kerry L. Cheesman
Biological Sciences
Capital University
Columbus, Ohio

Introduction

Multiple-choice exams are widely used in college science classrooms (as well as for laboratory quizzes and exams). Multiple-choice questions have many advantages—perhaps the most important is that they can be graded quickly and easily, and they can be graded by either human or machine. The “clicker” systems often used in large lecture rooms are well adapted for answering multiple-choice questions, and they can be used for “instant quizzes” with immediate feedback to students.

Instructor time is valuable, and in large classrooms the use of essay exams (the primary alternative) can quickly become overwhelming, causing students to wait for feedback for prolonged periods of time. Feedback on progress needs to be as rapid as possible, and essay questions do not lend themselves to that. Essay grading can also tend to be biased by any number of factors (time of day, personal biases, differences between graders, lack of openness to new interpretations, and so on).

Finally, most graduate entrance exams (including the GRE, MCAT, and DAT) are based on multiple-choice questions. Many later exams, such as the medical board exams, are also multiple-choice. Therefore, it is important to make sure that students are prepared for higher-order multiple-choice exams and the reasoning that is required to answer the questions in a proficient manner. Undergraduate science instructors can
help students be well prepared by using higher-order multiple-choice questions for assessment of course material starting in the freshman year.

Assessment must match one’s teaching style—inquiry teaching must be followed by assessment techniques that match the inquiry method of teaching. If one follows the learning cycle (5E or other similar models), assessment is encountered throughout the teaching and learning continuum, and that assessment must be related to the phase of the cycle (exploration, extension, etc.). Certainly higher-order questions capture the essence of exploration and extension much better than lower-order questions do.

The use of higher-order questions does not mean an end to using lower-order questions. Rather, we are referring to a shift from the typical 80–90% lower-order questions typically found in college science exams toward a balance between lower- and higher-order questions. The goal of undergraduate science instruction should be critical thinking rather than memorization. Many students come to the university with the assumption that science is just a lot of memorization, and college instructors often need to work hard to destroy that myth. However, that myth is often kept alive by the choice of questions used on the exams. If they favor knowledge-style questions, then students will continue to believe that science is mostly about memorization rather than about inquiry and analysis.

Understanding Bloom’s Taxonomy

Most college instructors are familiar, on some level, with Bloom’s taxonomy of learning (Bloom et al. 1956). Much has been written about the use of Bloom’s taxonomy in the construction of exam questions, but few instructors take to heart the need to use all of the levels instead of just the first two in constructing examination questions. Here is a quick review of Bloom’s taxonomy as it relates to the teaching of college science.

1. Knowledge: the ability to remember/recall previously learned material.
   
   Examples of behavioral verbs: list, name, identify, define, show
   
   Sample learning objectives in science: know common terms, know specific facts, know basic procedures and methods

2. Comprehension (understanding): the ability to grasp the meaning of material, and to explain or restate ideas.
   
   Examples of behavioral verbs: chart, compare, contrast, interpret, demonstrate
   
   Sample learning objectives in science: understand facts and principles, interpret charts and graphs, demonstrate laboratory methods and procedures

3. Application: the ability to use learned material in new situations.
   
   Examples of behavioral verbs: construct, manipulate, calculate, illustrate, solve
   
   Sample learning objectives in science: apply concepts and principles to new situations, apply theories to practical situations, construct graphs and charts

4. Analysis: the ability to separate material into component parts and show relationships between the parts.
   
   Examples of behavioral verbs: classify, categorize, organize, deduce, distinguish
   
   Sample learning objectives in science: distinguish between facts and inferences, evaluate the relevance of data, recognize unstated assumptions

5. Synthesis: the ability to put together separate ideas to form a new whole or establish new relationships.
   
   Examples of behavioral verbs: hypothesize, create, design, construct, plan
   
   Sample learning objectives in science: propose a plan for an experiment, formulate a new scheme for classifying, integrate multiple areas of learning into a plan to solve a problem
6. **Evaluation**: the ability to judge the worth or value of material against stated criteria.

*Examples of behavioral verbs*: evaluate, recommend, criticize, defend, justify

*Sample learning objective in science*: judge the way that conclusions are supported by the data

It is a common misconception that as one climbs the scale of Bloom’s taxonomy, the difficulty of the questions increases. The increase in cognitive demand associated with higher-order questions refers to the complexity of the questions, not the difficulty. Higher-order questions require a different set of cognitive demands, but they are not necessarily more difficult.

### Writing Multiple Choice Questions

Higher-order multiple-choice questions can be as easy or as difficult to construct as lower-order questions. Good-quality questions are essential to being able to truly assess a student’s knowledge and understanding of the subject matter in any area of science.

Before attempting to construct individual questions, think about the purpose of the questions. In general, the purpose should be to assess what students know and don’t know, and how students are able to construct knowledge based on prior learning. Therefore, avoid “trick” questions that may confuse students who understand the material. Avoid using prepared test banks written by the author of the textbook or other contracted writers. Honest assessment must match the teaching style employed, not the style of the textbook or the style of your colleagues. Note: You cannot ask higher-order questions if your teaching style mandates only recall.

Writing good multiple-choice questions takes time—a well-constructed test can’t be written in a single day. Questions need to be written, reviewed for clarity, and often revised. Questions need to be constructed in such a way that they neither reward test-wise students nor penalize those whose test-taking skills are less developed. The purpose is to assess student learning, and therefore each question needs to be clearly designed to achieve that goal. Remember that higher-order questions take longer to answer than recall questions, so plan accordingly in the construction of the test.

To construct a higher-order multiple-choice question, start by constructing the stem. The stem should pose a problem or state a question. Familiar forms include case study, premise and consequence, problem and solution, incomplete scenario, and analogy. The stem may involve pictures and diagrams or just words.

Write the stem as clearly and simply as possible. A student should be able to understand the problem without having to read it several times. Always try to state the problem in a positive form, as students often misread negatively phrased questions. Avoid extraneous language that is irrelevant to the question. While some authors believe this helps separate those who truly understand from those who don’t, too often it confuses even the well-prepared students, leading to unreliability of the question.

Never use double negatives. Avoid “which of these is the best choice” unless that format is integral to the learning objectives. Be sure to include in the stem any words that are redundant to all of the answers, and use “a(n)” or “a/an” to avoid eliminating any of the answers as mismatches.

Once the stem is constructed, proceed with writing the responses. Write the correct answer first. This allows you to be sure it is well constructed and accurate, and allows you to match the remaining answers to it. Avoid verbal cues, and certainly avoid lifting phrases directly from the text or class notes. Be sure that the incorrect responses match the correct one in length, complexity, phrasing, and style. For instance, in the following example, the mismatch of the answers makes it easy to guess the correct response even if one has little knowledge of the subject material.

The term “side effect of a drug”:

a. refers to any action of a drug in the body.
other than the one the doctor wanted the
drug to have
b. is the main effect of a drug
c. additionally benefits the individual

Distracters (incorrect answers) must be incor-
rect yet plausible. If a recognizable key word ap-
ppears in the correct answer, it should appear in
some of the distracters as well. Be sure to check—
will the answers help to distinguish why a student
got it wrong? This is an important part of assess-
ment that is often overlooked by instructors, but is
a critical part of helping students to learn.

Avoid microscopic distinctions between an-
swers, unless this is a significant objective of the
course. Be sure to stagger the correct responses in
their order (use all answer positions as equally as
possible). Limit the number of options—most au-
thors agree that 4–5 answers is plenty, and there is
no assessment advantage in using more than five.
Use all, always, never, none, etc., rarely. These are
answers that students have been programmed to

shy away from and may distort the question as a
valid assessment tool. Likewise, use all of the above
and none of the above sparingly.

When all exam questions have been con-
structed, check each one to see where it falls in
Bloom’s hierarchy. Construct a simple table such
as that shown in Table 1 to see the distribution of
questions. If the questions are disproportionately
distributed, then rewrite enough questions to bal-
ance the exam between lower-order and higher-
order questions.

**Examples of Multiple-Choice Questions at Each Level**

The following examples illustrate the construction
of multiple-choice questions that fit the higher
levels of Bloom’s taxonomy. For most an expla-
nation is included describing why it fits where it
does, and what a student needs to know to be able
to answer the question correctly.

**TABLE 1. Sample table showing the distribution of questions**

<table>
<thead>
<tr>
<th>Topics/Objectives</th>
<th>Recall (knowledge, comprehension)</th>
<th>Application (application, analysis)</th>
<th>Evaluation (synthesis, evaluation)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Application Questions

1. Susie and Bill are both healthy and have healthy parents. Each of them has a sister with **autosomal recessive cystic fibrosis**. If Susie and Phil have a child, what is the probability that it will be born with cystic fibrosis?
   a. 0
   b. 2/3
   c. 1/2
   d. 1/4
   e. 1/9
   (To answer this question correctly, one must understand the terms **autosomal** and **recessive**, and also understand the concepts of probability as applied to human genetics. In this question the student must apply those concepts to a family situation not studied before. The incorrect responses are constructed to find misconceptions/misunderstandings about genetic probabilities.)

2. A total of 100 students at Capital University were tested for blood type. The results showed 36 were type O, 28 were type A, 28 were type B, and 8 were type AB. The frequency of the A allele is therefore:
   a. 0.10
   b. 0.14
   c. 0.28
   d. 0.56
   e. 0.64
   (To answer correctly, a student must know the formula for allele frequency and be able to calculate it from the given data. The answers were chosen to help find misunderstandings about genetic probabilities.)

3. Evolutionary forces have produced an unusual plant, the Indian Pipe, that has no chlorophyll. Therefore, the plant must:
   a. make its own food
   b. absorb food made by other organisms
   c. photosynthesize without chlorophyll
   d. respire without taking in food
   e. use chlorophyll from other plants
   (To answer this question the role of chlorophyll in energy transformation must be understood. The student must apply the concepts of energy transformation/lack of chlorophyll to a logical new endpoint. All of the answers are plausible and help to distinguish where an understanding of energy transformation is incomplete.)

4. Which of the following compounds should have the highest boiling point?
   a. CH₃CH₁CH₂CH₃
   b. CH₅NH₂
   c. CH₃OH
   d. CH₃F₂
   (To answer correctly, a student must understand the concept of boiling point and the role of various constituent chemical groups in raising or lowering the boiling point.)

Analysis Questions

1. When a solid ball is added to a graduated cylinder with water in it, the water level rises from 20 ml to 50 ml. What is the volume of the ball?
   a. 20 ml
   b. 30 ml
   c. 50 ml
   d. 70 ml
   (In this example, a student must understand the concept of volume and not get distracted by the spherical nature of the added object. The answers are designed to give the instructor a sense of the misunderstanding of volume.)

2. From the graph shown here, determine when maximal carrying capacity has been reached.
   a. point A on the graph
   b. point B on the graph
   c. point C on the graph
   d. point D on the graph
   (To answer correctly a student must be able to see the relationships among the various organisms, interpret those relationships, and evaluate the ecosystem relative to the organisms shown on the graph.)
3. During an otherwise normal pregnancy, a woman begins to experience light-headedness and a decline in energy levels near the end of the first trimester. Which of the following is the most likely cause of her symptoms?
   a. lack of B vitamins due to poor diet
   b. decline of blood pH due to overuse of muscles
   c. decrease in blood pressure due to expanding fetal circulation
   d. decline in estrogen levels due to ovarian shutdown
(All of these answers involve factors that could cause tiredness in a woman. To determine the most likely cause in this scenario, a student needs to understand the basic mechanics of pregnancy and the biochemical changes that occur during it. The answer given shows a student’s ability to carefully analyze the situation and determine causality.)

4. The seeds of various plants vary in size from a fraction of a millimeter to several centimeters. The most critical factor controlling the size seed a plant produces is
   a. size of the maternal flower
   b. projected size of the animal pollinator
   c. quantity of the abiotic pollinator
   d. length of predicted dormancy
   e. method of distribution of seed
(To answer correctly, a student must put together knowledge of X and Y as compounds with knowledge of their dissociations and the reactions of the individual components. All of the answers reflect outcomes that the student has previously experienced when two compounds are mixed.)

**Synthesis Questions**
1. Domoic acid, produced by diatoms, has been found to bind to hippocampal glutamate receptors. If a person were to accidentally consume a lot of shellfish contaminated with this organism, what effect might be expected?
   a. blindness
   b. deafness
   c. amnesia
   d. aphasia
   e. rigidity
(To answer this question, one must understand the role of the hippocampus and the role of glutamate in this area of the brain. Here a student’s knowledge of the brain is used to establish a new relationship beyond those studied in class. All of the answers are logical outcomes of brain dysfunction and help the instructor to pinpoint the misunderstandings that students have.)

2. A neighbor found some mammal bones on the ground—a keeled sternum and an ulna, where the olecranon occupies 30% of the length of the bone. These bones most likely came from what type of mammal?
   a. flyer
   b. climber
   c. runner
   d. digger
   e. swimmer
(Here a student must understand the various bones and what their functions are. The student must then formulate a relationship between the type and formation of the bones and the activity that it would promote in a mammal.)

3. What would be the most logical result of mixing X and Y, both solubilized in distilled H₂O at room temperature?
   a. precipitation of a solid
   b. a change in color of the liquid
   c. a rapid rise in temperature
   d. a rapid decrease in temperature
(To answer correctly, one must understand the various bones and what their functions are. The student must then formulate a relationship between the type and formation of the bones and the activity that it would promote in a mammal.)

**Evaluation Questions**
1. Your fitness regimen involves jogging on the school track 2–3 miles per day with a friend. On a particular day, about 15 minutes into your jog, your friend suddenly pulls up and falls down, grasping her right calf in pain.
Conclusion

Multiple-choice questions may be used effectively to assess student learning as long as they are constructed properly and include assessment of higher-order thinking. Taking the time to construct good stems and good answers, not only on the exams but also in the daily questions posed during the lesson, is well worth the effort. Practicing coming up with conclusions for critical-thinking questions is as important in creating high-level thinking in students as designing good-quality multiple-choice exams. Asking only non-challenging thinking questions during class is a waste of both the instructor’s and students’ time and does little to help assess student learning.

Reference

As a “gateway” instructor for more than 30 years, I’ve learned a few things about assessing the “typical” community college student. “Gateway” at my institution is the polite euphemism for suggesting you’ll always be teaching the non-science majors with the slim hope that some may eventually learn to tolerate the subject. It’s been nearly 20 years since I’ve evaluated one true science discipline major among the hundreds of students in my classes each academic year. As STEM students are nonexistent in my classes, the best I can hope for is the integrated science major in education.

“Typical community college student” is an oxymoron. There isn’t one. Some of my college freshmen are older than I am, returning to college for an opportunity at a retirement career. A few freshmen are dual-enrolled high school students, 16 or 17 years old. Some of my students are parents of young children, and some are grandparents. Today’s community college students come from all walks of life and include working adults and recent immigrants or refugees from foreign lands. Included in this mixture is the university’s student of choice, whose parents chose the inexpensive route for their first two years of college.

So, how does a “gateway” instructor assess the learning of the “typical” community college student? Included in this chapter are 11 assessment lessons I’ve learned along the journey.
1. Assessments need to be frequent and scaffolded for legitimate success. Working adults don’t want only one or two assessments of their progress during the semester. They want to know exactly how well they are performing in the class at each assignment. They want feedback, personal and directed. These are practical folks! Not surprisingly, they want good grades, but they also want real success. They don’t mind being challenged, especially after they’ve been successful. Successful gateway instructors know that a 16-week semester might have seven or eight large exams. The first exam should be the least challenging. Each subsequent exam should be more challenging. Students want their efforts to show. They want to believe they’re progressing, that hard work pays off.

The experts might suggest that assessments be formative, giving frequent feedback toward the mastery of content. Classroom assessment techniques (CATS) are usually formative assessments, and might include the quick “think pair share” or the “one-minute paper” (Angelo and Cross 1993). Students want definite feedback on all formative assessments. They want to know that you’ve sincerely read each and every one. Formal grades aren’t required, just your attention and constructive remarks in some format.

Summative assessments are aligned with the evaluation of content mastery or the completion of instruction. Many community college students need state or national benchmarks (standards) for their instruction in their trade or vocational courses. My students see the state and national benchmarks for science education (Roseman and Koppal 2006), and they know that like other professionals, they need to meet those standards. Unlike their certification exams for careers, my class is only the beginning of their journey toward scientific literacy. Benchmarks and standards are a goal for attainment with the expected outcome of lifelong scientific literacy.

2. Assessments and exams are not always the same thing. Assessments come in many forms. In the assessment report I file each year to the divisional office to show that my nonscience majors course is worthy of the title of science “CORE” (which means it meets the criteria of inquiry, shows the processes and limitations of scientific thought, and analyzes data), no fewer than 12 different assessment techniques are listed.

Assessments include the exams, concept maps (two varieties), Vee diagrams, laboratory reports, and capstone projects. All of these give the instructor information about the students’ learning and their mastery of content. Alternative assessments that are real, targeted to the content can be more revealing than an exam. Anxiety plays a role in exam taking, but a student has control over a project. Presentations or projects that allow for research, sharing of ideas, and collaboration are valid assessments. These include contextual, problem-, case-, or performance-based assessments.

3. Embedded assessment across multiple sections has advantages and disadvantages. Community colleges are notorious for having large numbers of adjunct professors. Mine is not an exception. Subsequently, as the full-time professor responsible for reporting on multiple course offerings—even when I am not the “instructor of record”—my job becomes very challenging. Embedded questions on each exam allow for a logistically simple method for tracking all sections of a single course. Instructors simply provide me detail of the embedded questions after each exam. That’s the advantage. The disadvantage is that I have no idea why the students miss the embedded questions on particular topics. The variables are too numerous. I have a vague idea within my own classes, as I can monitor absences or recall the day in class when the topic was discussed. There is no information from the classes I didn’t teach. The number is a cold sta-
tistic without any qualitative information. After a period of time, the embedded questions on exams must change in their wording. Since the statistical report depends on data from the previous year, altering the question requires a whole set of rationale, without qualitative information. That’s another disadvantage.

4. Listening provides more assessment value than talking. As my students engage in their laboratory activities, I listen. I listen to their interactions, collaborations, and arguments. I walk to each lab table, and I listen. I can learn a tremendous amount about their learning and their assimilation of the content by listening. During the course of the semester, they become very accustomed to me walking to each table without saying anything as they work. I learn a lot about their thinking processes by listening. Each unit of instruction also begins with me listening. Each collaborative group is asked to list prior knowledge about topics in the unit. Together we summarize. We use the prior knowledge expressed in discussions to increase the depth of knowledge on the topics. Pre-assessment starts with listening.

Each unit of the courses I teach starts with a series of connection questions. What do you already know about the topic? How can learning this information be useful to you? What are you looking forward to learning about this topic? Each unit also ends with reflection. What did you learn about this topic? Did anything you learned in this unit change your mind? How will this information be useful for your future? Postassessment also starts with listening. I’ve learned that students do not necessarily answer these questions unless they are explicitly asked. If there exists a possibility that you could be called upon and expected to respond directly to a specific question, you prepare a response. Without that potential accountability, it’s a rare student who prepares a response or who is introspective without the prompting questions.

5. A wrong answer has tremendous value. A well-thought-out, detailed wrong answer gives you lots of information. It provides you opportunity to correct a misconception or to craft a discrepant event to allow the learner to construct a more scientifically accurate response. Providing the question in advance and allowing two minutes to think before calling on a student to respond yields more information than simply calling on a student. Wait time also works well (Rowe 2003). Giving students a 30-second warning before expecting a response is very powerful. “I’m going to ask <student’s name> to respond to the next question” gives that individual a few extra seconds to compose an answer. The responses are more complete, even when they are wrong.

6. Assessment need to be clearly tied to outcomes, objectives, or learning targets. Both instructor and student need to clearly know the purpose of the assessment. What are we evaluating? Communications of expectations are important. Providing the format of the exam gives students an opportunity to prepare appropriately. You study differently for a written essay than a multiple-choice exam. You prepare differently for a presentation than for a discussion.

7. Assessments that are viewed as “products” by students are sources of pride. I have many “product” assessments in my classes. I’m always surprised by college students who have told me that their perfect score concept map was hung on their refrigerator! Or the lab report with the phrase “Well done!” was read over the dinner table. It seems it doesn’t really matter how old we are, a well-done product is a source of pride. Community college students know the rewards of hard work. They work tremendously hard in the challenges of everyday life. They can do exceptional work when the assessment is viewed as a product.

8. Detailed constructive feedback on assessments is essential. It takes about six hours to correct a stack of 24 lab reports, if they are well writ-
A set of poorly written lab reports takes about twice as long. Each report needs carefully worded constructive feedback. Rubrics on written assignments (and oral for presentations) are given in advance, and students are expected to follow the same criteria for excellence in writing as they would in a composition class. Even on exams, common mistakes are explained. The feedback on the exam is another opportunity to teach.

Owning and expressing your expectations for their success is crucial. Students rise to the challenge of high expectations. When given rationale for a challenge, they accept. They will even accept the frustrations of disequilibrium if they understand the rationale. Explicit reasons for content expectations are essential. Community college students will accept the “because it’s on the test” but are likely to ask you why it’s on the test. They want a more practical reason for learning the content. Ideally, the reason is tied to a potential career, or an everyday application.

Assessments need to be varied, perceived as fair and attainable, and evaluated both objectively and subjectively (Mintzes, Wandersee, and Nowak, 1999). The brain loves novelty. It fatigues when offered routine. With 12 different types of assessments throughout the semester, fatigue is more physical than cerebral! Each exam has a variety of question types. Students create or correct concept maps; they evaluate true and false statements, correcting the false. They also write brief answers and traditional multiple paragraph essays. Each exam also has multiple-choice questions and paragraph completions. Students analyze their exam results at the conclusion of each unit and write goals to improve weak performance areas. Students know how exams are evaluated. They know that each section is evaluated independently without my knowledge of the test author. They also know when the sections of the exam are totaled, I often write encouraging remarks on their progress (e.g., “Nice improvement on this multiple-choice section; keep working!”) Statistical analysis is given on the entire class performance, and the class discusses improvement strategies for the next exam. Besides the exams, alternative assessments are a near daily occurrence.

Ideally, assessments inform teaching, and self-assessments can even inform the learner. Assessment is not only about evaluating the learning process. It should change the teaching process. Each assessment should inform the instructor as to needed changes in pedagogy, presentation or missing fundamentals for conceptual understanding. Self-assessments can provide the learner with great potential to change.

The view from the “gate” as I encourage students to consider the science disciplines is generally positive. Together, the students and I investigate, listen to each other, and plan our journey together. Although assessment reports are needed for multiple levels (divisional, program, departmental) there is enough consistency across the requirements that only the perspective changes. Not only have I learned how to assess my “typical” community college students so that I know what they are learning, I’ve also learned how to teach better science through our shared assessments.

References


Roseman, J. E., and M. Koppal. 2006. Ensuring that college graduates are science literate:

Index

Page numbers in boldface type refer to figures or tables.

A
AAHE. See American Association for Higher Education
Administrative support for system, institutions using resources, 9
ADT. See Astronomy Diagnostic Test
Advanced Placement examinations, 21
American Association for Higher Education, 9
The American Biology Teacher, 115, 117
Analysis, Bloom’s taxonomy, 36, 50
Analysis questions to assess higher-order thinking, 39–40
AP examinations. See Advanced Placement examinations
Application, Bloom’s taxonomy, 36, 50
Application questions to assess higher-order thinking, 39
Appropriate Workload Scale, 132
Armed Services Vocational Aptitude Battery, 21
Assessments
college science laboratory, alternative forms, 49–54
defined, 7
with e-mail, 87–88
electronic portfolios for, 59–66
faculty-owned, building culture of, 7–13
5-E lesson plan development, 81–84
higher-order thinking, multiple-choice questions, 35–41
to link research to practice, 31–32
for motivating students, 151–153
multiple-choice, 43–48
nature of science, alternate form of assessment, 77–79
patterns, 129–136
peer, 89–93
performance indicator development, 129–136
practices jeopardizing, 137–145
quantitative, in large undergraduate science classrooms, 15–30
with student remotes, 87–88
through creation of scientific posters, 55–57
varied assessment, 147–149
wetscience, 67–75
Astronomy Diagnostic Test, 18, 22
ASVAB. See Armed Services Vocational Aptitude Battery

B
Backwash effect, 129
BCI. See Biology Concept Inventory
Benchmarks for Science Literacy, 59
Bio 2010, 59
Biology Concept Inventory, 19
Biweekly exams, 151
Bloom’s taxonomy, 36–37, 50–51, 109
Book reviews, student-authored, 115–116

C
Challenges in building culture of faculty-owned assessment, 8
Child, explaining to, as alternative assessment, 53
Classification as alternative assessment, 52
Classroom methodologies, quantitative assessment, 16–20
Classroom-tested practices, instructions, implementation, 85–120
Clear Goals and Standards Scale, 132
CMS. See Course management systems
College classroom assessment, 33–85
alternative assessments, 49–66
teacher education, 67–84
traditional assessments, 35–48
College science laboratory, alternative forms of assessment, 49–54
Bloom’s taxonomy, 50–51
types of assessment, 51–52
Commission on Future of Higher Education, 8
Communication, in culture of faculty-owned assessment, 10
Completion of exam evaluation over course of semester, question regarding, 100
question regarding, 101
Comprehension, Bloom’s taxonomy, 36, 50
Construct validity
defined, 3–4
in validation of survey instruments, 4
Constructive alignment, 129
Content validity, defined, 3–4
Course management systems, assessing learning via, 22–25
Create, posters, 52
Criterion-related validity, defined, 3–4
Culture of faculty-owned assessment, 7–13
American Association for Higher Education, 9
assessment, defined, 7
building, obstacles to, 8
challenges, 8
Commission on Future of Higher Education, 8
communication, 10
Faculty Learning Communities, 11
faculty time, 8
foundations, 8
isolation, 8
No Child Left Behind Act, 8
planning, 9–10
results, 12–13
skepticism, 8
values
clarification of, 9
structure reflecting, 10–11
varying involvement, 8
visible, vocal leadership, 11–12

D
DAT. See Graduate entrance exams
Data-oral alternative assessment, 53
Design/build/produce as alternative assessment, 52
Diagnostic learning log, 95–97, 96
Distracters, multiple-choice assessments, 46, 46
DLL. See Diagnostic Learning Log
Documenting learning outcomes, multiple measures, 9
“Donations Please” project, 91

E
e-mail
formative assessments using, 87–88
sending grades via, 153
e-portfolios. See Electronic portfolios
Edmiston, Dr. Judith, 109
Electronic portfolios, college science course assessment, 59–66
assessing, 61–63
checklist for, 62
defined, 60
National Coalition for Electronic Portfolio Research, 60
qualitative scoring, 63
rubric for, 63
Embedded assessment, advantages, disadvantages, 124–125
Engagement, student, 95–97
Epistemological commitments underlying activities of science, 79
Evaluation, Bloom’s taxonomy, 37, 51
Evaluation questions to assess higher-order thinking, 40–41
Exam analysis, instructor perspective, 109–110
Exam corrections, analysis, student perspective, 107–108
Expert-teaching others in class, as alternative assessment, 53–54
Explaining to child, as alternative assessment, 53

F
Face validity, defined, 3–4
Faculty Learning Communities, 11
Faculty-owned assessment, culture of, 7–13
American Association for Higher Education, 9
assessment, defined, 7
challenges, 8
Commission on Future of Higher Education, 8
communication, 10
Faculty Learning Communities, 11
faculty time, 8
foundations, 8
isolation, 8
No Child Left Behind Act, 8
planning, 9–10
results, 12–13
skepticism, 8
values
clarification of, 9
structure reflecting, 10–11
varying involvement, 8
visible, vocal leadership, 11–12
Faculty time as obstacle to building campuswide assessment program, 8
Failure prediction, assessment as tool for, 25–26
FCI. See Force Concept Inventory
Field-testing multiple-choice assessments, 47
5-E lesson plan development, science content learning with preservice teachers, 81–84
FLCs. See Faculty Learning Communities
FMCI. See Force and Motion Concept Evaluation
Force and Motion Concept Evaluation, 21–22
Force Concept Inventory, 18, 22, 25
Foundations of culture of faculty-owned
assessment, 8
Future classes, using exam evaluation in, question regarding, 101

G
Gateway instructor, assessment lessons learned as, 123–127
“Gateway” instructor, assessments learned as, 123–127
General assessment topics, 1–32
Good Teaching Scale, 132
Graduate entrance exams, higher-order thinking, 35
GRE. See Graduate entrance exams
Group assessments, as alternative assessment, 54

H
Helpful of exam evaluation, question regarding, 101
Higher-order questions, multiple-choice tests using, as alternative assessment, 54
Higher-order thinking writing/using multiple-choice questions to assess, 35–41
analysis questions, 39–40
application questions, 39
Bloom’s taxonomy, 36–37
distribution of questions, 38
evaluation questions, 40–41
graduate entrance exams, 35
multiple-choice questions, 37–41
synthesis questions, 40
Hypothesis modification activity, 103–106
evaluation, 105–106
implementation, 103–105
student agreement on evaluation, 105

I
In-groups, out-groups, comparison of learning between, 20–22
Inquiry-based labs, scientific report, 111–114
Isolation, as obstacle to building culture of faculty-owned assessment, 8

J
Journal of College Science Teaching, 115, 117

K
Knowledge, Bloom’s taxonomy, 36, 50

L
Lab exams, 51
Lab reports, 51
Laboratory investigation scientific report, grading criteria, 112–114
Large undergraduate science classrooms, quantitative assessment, 15–30
failure prediction, assessment as tool for, 25–26
in-groups, out-groups, comparison of learning between, 20–22
pretest scores, final grade correlations, 25
“strugglers,” capturing, 26–27
Law School Admissions Test, 21
Leadership, in culture of faculty-owned assessment, 11–12
Linking research, practice, in organizing course on student assessment, 31–32
Listening to students, importance of, 4–5, 125
LSAT. See Law School Admissions Test

M
MCAT. See Medical College Admissions Test
Medical College Admissions Test, 21, 35
Motivation, 101, 151–153
actively proctor exam, 152
class time, 152
e-mail list, sending grades over, 153
fairness of tests, 152
frequent exams, 151
promoting class work, 152
relating current topics to past topics, 152
rewarding students, 152
short exams, 151
source of questions, mentioning in class, 152
Multiple-choice assessments, 45–48
analysis questions, 39–40
application questions, 39
to assess higher-order thinking, 35–41
Bloom’s taxonomy, 36–37
distracters, 46
distribution of questions, 38
evaluation questions, 40–41
examples, 38–41
field-testing, 47
graduate entrance exams, 35
scoring guide, 47
synthesis questions, 40
teach students to take, 43–44
using higher-order questions, as alternative assessment, 54
writing, 37–38

N
National Coalition for Electronic Portfolio Research, 60
National Science Education Standards, 59
National Science Foundation, established teaching fellowships, 73
Nature of science, gauging, 77–79
The New York Times Book Review, 115
No Child Left Behind Act, 8
Nonscience major, explaining to, as alternative assessment, 53

O
Open-ended written questions, 4
Organizing course on student assessment, practice, research, linking, 31–32
Outreach program, math, science, technology incorporation into service learning. See Wetscience

P
Patterns in assessment, 129–136
Appropriate Workload Scale, 132
backwash effect, 129
changing, 134–136
Clear Goals and Standards Scale, 132
constructive alignment, 129
evaluation tool, 132–134
Good Teaching Scale, 132
influence of teaching on assessment, 132
integrating assessment items into, 131–132
judging assessment pattern, 130–131
measuring, 134
student experience, assessment’s impact on, 130
Peer assessment, 89–93
Planning, in culture of faculty-owned assessment, 9–10
Posters, 55–57
as alternative assessment, 52
assessment of students’ learning through, 55–57
scoresheet criteria, 57, 57
Practice, linking research, in organizing course on student assessment, 31–32
Practices for improving assessment, 121–153
Practices jeopardizing student assessment, 137–145
“Preparing Tomorrow’s Teachers to Use Technology,” 72
Pretest scores, final grade correlations, 25
Principles of survey instrument development, 4
PT3. See “Preparing Tomorrow’s Teachers to Use Technology”
Putting assessment findings into practice, 9

Q
Quizzes, weekly, 51

R
Reflection on exams, promotion of, 99–101
completion of exam evaluation, question regarding, 100–101
future classes, using exam evaluation in, question regarding, 101
helpful of exam evaluation, question regarding, 101
motivation to complete exam evaluation, question regarding, 101
student survey results, 100–101
Remote student, formative assessment with, 87–88
Rubrics, classroom-tested, implementation, 85–120

S
SALG. See Student Assessment of Learning Gains
SAT. See Scholastic Aptitude Test
Scaffolding, 124
Scholastic Aptitude Test, 21
Science for All Americans, 59
Science News, 115
Scientific American, 115
Scientific posters, 55–57
scoresheet criteria, 57
Scientific report, inquiry-based labs, 111–114
Scientific Report of Laboratory Investigation Grading Criteria, 111–114
Show/demonstrate as alternative assessment, 52
Skepticism, as obstacles to building campuswide assessment program, 8
Staff, work on, assessment as, 9
Statistical Reasoning Assessment, 19
Structure reflecting values, in building culture of faculty-owned assessment, 10–11
“Strugglers,” capturing, 26–27
Student Assessment of Learning Gains, 17
Student experience, assessment’s impact on, 130
Student-led teaching models, 117–120
grading criteria, 118–120
Student remotes, formative assessment with, 87–88
Student survey results, 100–101
Students performing exam corrections, outline for, 107–108
Success, supporting, formative activity focused on, 9
Survey instrument validation, 3–6
construct validity, 4
definitions, 3–4
listening to students, 4–5
open-ended written questions, 4
principles of survey instrument development, 4
Synthesis
Bloom’s taxonomy, 36, 50
questions to assess higher-order thinking, 40

T
Taxonomy, Bloom’s, 36–37, 50–51, 109
Technical support for system, institutions using resources, 9
Third International Mathematics and Science Study, 73
TIMSS. See Third International Mathematics and Science Study
Topics in assessment, 1–32
Transparent assessment, 9

V
Validation of survey instruments, 3–6
construct validity, 4
definitions, 3–4
listening to students, 4–5
open-ended written questions, 4
principles of survey instrument development, 4
Validity, defined, 3–4
Values, clarification in building culture of faculty-owned assessment, 9
Varied assessment, 147–149
dynamic vs. static, 148
individual vs. group performance, 148
individual vs. group review, 148
product vs. process, 148
Varying involvement, as obstacle to building culture of faculty-owned assessment, 8
Views of Nature of Science Questionnaire, 78
typical questions from, 78
Visible, vocal leadership, in culture of faculty-owned assessment, 11–12
VNOS. See Views of Nature of Science Questionnaire

W
Weekly exams, quizzes, 151
Wetscience, math, science, technology incorporation into service learning outreach program, 67–75
Writing/using multiple-choice questions to assess higher-order thinking, 35–41
analysis questions, 39–40
application questions, 39
Bloom’s taxonomy, 36–37
distribution of questions, 38
evaluation questions, 40–41
graduate entrance exams, 35
multiple-choice questions, 37–41
synthesis questions, 40
Written explanation alternative assessment, 53
Written questions, open-ended, 4