

How is Science Learning Assessed at the Postsecondary Level? Assessment and Grading Practices in College Biology, Chemistry and Physics

Karleen Goubeaud

Published online: 1 October 2009
© Springer Science+Business Media, LLC 2009

Abstract The role of assessment in higher education is gaining importance as accountability requirements intensify and as assessments are increasingly recognized as having potential to improve teaching and learning. During the last two decades, educators have begun implementing a wider variety of assessment types including alternative and student-centered assessment practices. However, few research studies have examined the extent that college science faculty use such practices. This large-scale descriptive study utilized a nationally representative sample of higher education faculty from the US Department of Education to examine the assessment and grading practices of college science faculty from 2 and 4-year higher education institutions. When data was disaggregated by science discipline, statistically significant differences were found among physics, chemistry and biology faculty's assessment and grading practices. Biology faculty used a larger repertoire of assessment types overall, and used assessments that have potential to enhance the learning process more than chemistry and physics faculty. Physics and chemistry faculty graded on a curve more often and used competency-grading practices less often than biology faculty. Assessment practices that could be considered formative with potential to promote student learning appear to be underutilized by all science faculty.

Keywords Science education · Postsecondary education · Classroom assessment · Biology · Chemistry · Physics

Introduction

The critical role of assessment in education has been underscored by advances in cognitive science that enhance our understanding of how students learn. Research over the last several decades indicates that students learn when they construct their own understanding as a result of active engagement in meaningful learning experiences. In contrast, traditional approaches to learning view students as passive receivers of information (Yager 1991). This difference in viewing both the teacher and student as actively engaged in the learning process necessitates new ways of conceptualizing both instruction and assessment. The shift in thinking about how students learn elevates assessment beyond a means of evaluation to an important component of the learning process, a change that requires that schools implement new assessment practices designed to enhance learning potential (Shepard 2000).

From a public and political perspective, assessment has become prominent as the accountability movement grows (Linn et al. 2002; US Department of Education 2001). At the higher education level, there is public concern that colleges and universities not only provide a quality education, but also produce evidence that demonstrates student competency. The public wants to ensure the quality preparation of science majors at the undergraduate level to increase persistence through graduation to prevent shortages of scientists and engineers in the US workforce (Butz et al. 2003) and to prepare future teachers who are competent in science content areas (Black 2003).

Assessments have potential not only to enhance the learning process and inform instruction but also to provide evidence to stakeholders that students are competent. Evidence for science learning can come in many forms depending on the type of learning being assessed.

K. Goubeaud (✉)
Department of Curriculum and Instruction,
Long Island University, C.W. Post Campus,
720 Northern Blvd., Brookville, NY 11548, USA
e-mail: karleen.goubeaud@liu.edu

Traditionally, educators viewed assessment as occurring after the lesson or unit to assess students' content knowledge. In contrast, alternative types of assessment can be implemented at any time during the learning process and can provide a broader range of options to capture students' progress towards a variety of learning goals. To account for these important types of student learning, a wide array of assessments matching curricular goals in science education are needed (Atkin and Black 2003).

Assessment Practices that Measure a Broad Spectrum of Learning Goals

Although almost all types of assessment are potentially valuable tools in science education, particular characteristics of assessments make them more suitable for measuring specific learning outcomes. Assessments that contain open-ended items provide opportunity for students to elaborate on their scientific understanding, making cognitive processes of students more visible. Through responses to open-ended items, students demonstrate their writing skill using the language of science. Students' elaborated writing responses can inform teachers of students' views of the nature of science, level of scientific literacy, and ability to interpret scientific evidence. The value of open-ended assessments has become increasingly more apparent as educators recognize the value of students' argumentation (Jiménez-Alexandre and Erduran 2008). Rich data from students' evaluation of scientific evidence is not as easily obtained from short-answer or forced-choice responses traditionally used in classrooms.

An assessment's open-endedness enhances its potential to measure science learning goals that ask students to demonstrate their learning in context. Students have opportunity to express their understanding in an authentic form that parallels the context of how the understandings would be used in real life (Wiggins 1998). Gronlund (2005) describes a continuum of open-endedness on which assessments can be ranked according to the degree students have opportunity to structure their response. Assessments that provide greater flexibility can put learning in context, be more authentic, and measure important facets of inquiry learning, rather than function purely as an evaluative tool (Bass and Glaser 2004).

On the other hand, if an assessment's purpose is to assess students' ability to distinguish between scientifically accurate views and misconceptions, then assessments utilizing forced-choice responses such as multiple-choice items might be a preferred item option. Carefully written stems that target key ideas or connections between ideas can be combined with distracters that represent plausible yet incorrect responses. Strategically-chosen distracters are key to assessing students' ability to discern errors from

accurate scientific statements. If students' responses are analyzed systematically, results from such multiple-choice assessment items can provide specific insight into student thinking to inform the learning process. In the last two decades, there has been a growing awareness of this issue among postsecondary science educators and efforts to create multiple-choice tests that would probe students' understanding more deeply and inform the instructional process. Physics Education Research literature (see studies referenced in McDermott and Redish 1999) demonstrates the increased use of high quality multiple-choice tests as diagnostic learning tools. For example, in the area of mechanics, several assessment instruments have been designed for this purpose such as the Mechanics Baseline Test (MBT) (Hestenes and Wells 1992) and subsequent development of the Force Concept Inventory (FCI) to assess conceptual understanding of Newtonian mechanics (Mazur 1997). These tools demonstrate a recent shift in college science faculty's use of assessment for improving student learning.

More recently, a project being conducted at Michigan State University is designing high-quality multiple-choice items and assessments that maximize learning potential (Richmond et al. 2008). The items are being refined through a rigorous research process, so that items have strong formative potential for understanding students' thinking. Although these studies and instruments are being developed and some are available to college science faculty, more evidence is needed to determine if these types of high quality science test items are being utilized to a significant degree in day-to-day college science teaching.

Assessment Practices for Formative and Summative Purposes

Gathering many different types of data to inform instruction for formative purposes as well as adequately grading and reporting for summative purposes is necessary during the instructional process. Qualitative and quantitative data each serve their purpose to provide teachers with information to identify students' strengths and weaknesses and to give to students meaningful, specific feedback to improve their learning performances. Sato et al. (2008) identified six dimensions of formative assessment that include a focus on teachers' ability to use a range of assessments to strategically promote students' learning.

The potential of a formative assessment to contribute to students' scientific inquiry learning not only depends on the assessment's design but also on how the assessment process is implemented (Ruiz-Primo and Furtak 2007). Traditionally, the process of "grading" has referred to checking students' work for errors, inaccuracies, or omissions before assigning a score or grade which is entered into the students'

school record. In contrast, grading in the context of formative assessment lends itself to a process where students are active participants. Students engage in a collaborative process where feedback is available from multiple sources such as peers and teacher (Chappuis and Stiggins 2002), allowing students' specific strengths and weaknesses in skills or understanding to be identified. For example, The Peer Instruction (PI) approach recommended by Mazur (1997) provides a model of involving students in the learning process in a formative assessment capacity where peers express and critique each other's scientific ideas during a lesson. Through the process, students recognize errors, misconceptions, or incomplete knowledge and become explicitly aware of their new understandings. Using this information students can be given opportunities to revise or connect new learning to previous knowledge as they improve learning performances and polish products. The new draft or revised product might reflect specific feedback from peers, teachers, and student self evaluation (Dochy et al. 1999). The effectiveness of the process depends on the types of data collected and how the data is used to modify classroom instruction (Black and Wiliam 1998; Sato et al. 2008).

For summative purposes, assessments are chosen based on both the function of the assessment and the conclusions to be drawn from the data gathered. Assuming an assessment is a valid measure of student learning a particular area, the grade on the assessment would indicate the degree of content mastery or the extent that students understand or have knowledge of that area. Criterion-referenced assessments reflect this purpose because they indicate the level of competence students have achieved, directly linking learning accomplishment to assessment scores. For this reason, many educational assessment experts prefer criterion-referenced tests in classrooms over norm-reference tests which adjust scores based on student comparisons (Gronlund 2005; Sadler 2005). Currently, standards are serving as the criteria for many assessments as schools at all levels are aligning these with curriculum. However, criteria-based grading which predates the standards movement in education can include any learning goal that teachers utilize to guide learning and assessment, acknowledging the varying degree that sets of standards are adequate to capture the most important science learning. While grading on a curve or other norm-referenced grading practices can be useful for some evaluation purposes, particularly school-level comparisons, they are less useful for classroom assessment, which should focus on the extent to which learning objectives have been met (Popham 2003).

Purpose of the Study

A variety, rather than a narrow repertoire of assessments, is necessary to be able to assess the skills, knowledge and competencies that students should demonstrate in college

science. An assortment of assessment types is recommended by the National Science Education Standards to assess the variety of types of student learning, stating that “all aspects of science achievement—ability to inquire, scientific understanding of the natural world, understanding of the nature and utility of science—are measured using multiple methods such as performance and portfolios, as well as conventional paper-and-pencil tests” (NRC 1996, p. 76).

Fortunately, more educators are becoming convinced of the value of formative assessment to enhance learning (Maclellan 2004). The use of a wider variety of assessments including alternative assessment appears to be increasing at the K–12 education level (Crane and Winterbottom 2008; Stiggins 1991; Ruiz-Primo et al. 2004). However, it is not clear from the literature the extent to which college science faculty implement various types of assessments at the college level. Few publications in the literature examine the assessment practices of higher education faculty. There are isolated examples of college faculty describing particular performance assessments used in their science classes. For example, Robyt and White (1990) documented the use of laboratory practical formats in biology and chemistry. Lab assessments could include a variety of item types including more open-ended questions where students explain their reasoning as well as short answer or selected response. Slater (1997) implemented portfolios in physics. However, of the studies examining science faculty's assessment practices published in the science education literature, most have small samples sizes or are limited to science courses at a single institution. Large-scale research is needed to examine how college faculty assess science learning and to measure the extent that a variety of alternative as well as traditional assessment practices are being used.

To address these limitations, the present study utilized a nationally representative sample of higher education faculty to examine the types of assessments being used in college science in the US. The purpose of the study was to (a) describe the assessment and grading practices of biology, chemistry, and physics faculty, and (b) compare the assessment practices of faculty from these science disciplines.

Method

Data Source

A sample of science faculty was drawn from the National Study of Postsecondary Faculty (NSOPF), a National Center for Education Statistics (NCES) dataset sponsored by the US Department of Education. This data set is the largest database of higher education faculty in existence and contains survey information from 28,576 higher

education faculty. The sample included faculty from all types of institutions, both public and private, and is representative of the composition of faculty in the United States in terms of demographics and other characteristics. Faculty who taught in undergraduate institutions in the subject areas of biology, chemistry and physics were drawn from the sample totaling 2,750 science faculty used for the present study.

Data Collection

The data were collected in a cross sectional study in 1999 by the NCES. A clustered, stratified sampling design was utilized. First, higher education institutions were selected proportional to the estimated number of faculty in each Carnegie Classification stratum. Second, the sample of faculty was clustered within the 819 institutions selected. Faculty were given the option of either completing a paper questionnaire returned by mail or by completing the questionnaire through the Internet. Participation in the NSOPF study was voluntary and coordinated through the higher education institution in which they taught. Additional technical details related to sampling and data collection can be obtained from the National Study of Postsecondary Faculty: 1999 Methodology Report (US Department of Education 2002).

Faculty were asked in a self-report questionnaire about the frequency they used particular types of assessment practices in their undergraduate classes. Faculty were asked “In how many of the undergraduate courses that you taught for credit during the 1998 Fall term did you use...” assessment types such as “multiple-choice exams,” “essay exams,” “short-answer exams,” “term/research papers”), formative assessment practices (i.e., “student evaluation of each others’ work,” and using “multiple drafts of written work”), and grading practices (i.e., “grading on a curve” and “competency-based grading”). Faculty reported whether they used these strategies in “all,” “some” or “none” of their classes.

Data Analysis

Descriptive statistics were used to report the proportion of faculty who implemented various types of assessment strategies in their classes. Chi-square analysis was conducted to compare assessment practices of science faculty in the fields of biology, chemistry, and physics and examine difference in their use of these assessments. All analyses were conducted with appropriate weights for the complex survey design of NSOPF:99 to adjust for differential probabilities of selection and nonresponse at the institution and faculty levels (US Department of Education 2002).

Results

The data indicated that there were significant differences among the assessment practices of biology, chemistry and physics faculty at undergraduate higher education institutions.

Assessment Practices of Science Learning

Multiple-Choice Exams

Multiple-choice items, due to their structure, measure students’ understanding by limiting responses selected by the test designer. In contrast to open-ended items, these items often measure lower cognitive levels of thinking, although the quality of items is dependent on their specific design and use (Martinez 1999). A greater proportion of biology faculty used multiple-choice exams than chemistry or physics faculty. For example, 39.0% of biology faculty used multiple-choice exams in “all” of their classes, compared with 21.3% of chemistry faculty and 17.9% of physics faculty. The majority of biology faculty (73.2%) used multiple-choice items in “some” or “all” of their classes collapsing these categories, compared to about half of chemistry faculty (56.4%) and slightly fewer physics faculty (44.8%). Differences between biology, chemistry, and physics faculty in their use of multiple-choice tests were statistically significant, $\chi^2(4, N = 2,754) = 219.940, P = .000$. See Table 1 for percentages of faculty who used various types of assessments and for results of the Chi square analysis.

Short-Answer Exams

Short-answer items, recently termed constructed-response, give test-takers more flexibility in expressing their response than multiple-choice items but less than extended-response items such as essays or research papers. If carefully designed these items can be a valuable and valid tool for assessing learning (Hogan and Murphy 2007). Over half of science faculty used short answer exams in “some” or “all” of their classes. Somewhat more chemistry faculty used short answer exams than either biology or physics faculty. For example, 75.3% of chemistry faculty used these items compared to 60.9% of biology and 62.7% of physics faculty. The differences among science faculty disciplines were statistically significant, $\chi^2(4, N = 2,754) = 32.692, P = .000$. See Table 1.

Essay Exams

Essay exams are open-ended assessment items that allow students to demonstrate their scientific understanding through elaborated narratives, sometimes called extended-response

Table 1 Types of classroom assessments used by science faculty

	Biology		Chemistry		Physics		χ^2
	N	%	N	%	N	%	
Multiple-choice exams							
All	599	39.0	86	21.3	146	17.9	219.94**
Some	525	34.2	142	35.1	219	26.9	
None	411	26.8	176	43.6	450	55.2	
Short-answer exams							
All	300	19.5	109	27.0	189	23.2	32.69**
Some	635	41.4	195	48.3	322	39.5	
None	600	39.1	100	24.8	304	37.3	
Essay exams							
All	296	19.3	80	19.8	161	19.7	11.19*
Some	504	32.8	110	27.2	221	27.1	
None	735	47.9	215	53.1	434	53.2	
Term or research papers							
All	355	23.1	59	14.6	134	16.4	55.63**
Some	549	35.8	108	26.7	256	31.3	
None	631	41.1	237	58.7	427	52.3	

* $P < 0.05$, ** $P < 0.001$

items. Jacobs (1992) recommends essay writing in higher education because it is better suited to assess complex learning outcomes than test items requiring students to merely recognize correct answers. About half of science faculty used essay items in at least “some” of their courses (biology, 52.1%; chemistry, 47.0%, physics, 46.8%). There were small but statistically significant differences among science disciplines: $\chi^2(4, N = 2,756) = 11.192, P = .000$. See Table 1.

Research and Term Papers

Open-ended assessments involving student writing such as term or research papers can facilitate conceptual change (Fellows 1994). Writing assessments are valuable tools for inquiry because students not only explain their ideas about scientific phenomenon, but also describe the reasoning and evidence that support their explanations (Takao and Kelly 2003). Less than half of chemistry and physics faculty assign term or research papers (41.3% and 47.7%) but somewhat more biology faculty (58.9%) use this strategy in “some” or “all” of their classes. The differences among science faculty disciplines were statistically significant, $\chi^2(4, N = 2,756) = 55.629, P = .000$. See Table 1.

Grading Practices for Formative Purposes

Students’ Evaluation of Each Others’ Work

In this type of assessment, students are engaged in some aspect of critically examining the quality of another

Table 2 Types of grading practices used by science faculty

	Biology		Chemistry		Physics		χ^2
	N	%	N	%	N	%	
Students’ evaluation of other students’ work							
All	193	12.6	22	5.4	107	13.1	70.61**
Some	442	28.8	67	16.6	153	18.7	
None	900	58.6	315	78.0	557	68.2	
Multiple drafts of written work							
All	156	10.2	15	3.7	74	9.1	30.13**
Some	342	22.3	68	16.8	146	17.9	
None	1036	67.5	322	79.5	596	73.0	

* $P < 0.05$, ** $P < 0.001$

student’s work. Involving students in the assessment process can facilitate learning. Feedback from peers helps students revise their work and broadens their thinking to understand others’ viewpoints. Students build their own understanding as they interact with one another, a process best implemented under the guidance of a content expert or instructor who monitors student discussions and facilitates learning in this context. Palomba (1999) recommends this practice in higher education and highlights the value of students assisting in grading or critiquing their peers’ projects or presentations in the present study. However, the majority of science faculty from all disciplines did *not* use peer assessment in their classes. About three-fourths of chemistry faculty (78.0%) did not use this type of assessment, followed by physics and biology faculty (68.2% and 58.6%). The differences among science faculty disciplines were statistically significant, $\chi^2(4, N = 2,756) = 70.611, P = .000$. See Table 2.

Multiple Drafts of Written Work

Unlike traditional assessments that are administered once after a unit or semester of study, alternative assessments often result in learning products that can be revised over time. Such assessments allow the purposes of assessment to be expanded from merely evaluating students’ performance to enhancing the learning process. In this context, giving students opportunities to produce multiple-drafts of written work becomes not only an assessment strategy but also an integral part of learning. Feedback from teacher, peers and student self-evaluation can be utilized to help students reflect on and improve the quality of their work. Less than one-third of science faculty from any science discipline asked students to produce multiple-drafts of written work. For example, 32.5% of biology faculty, 20.5% of chemistry faculty, and 27.0% of physics faculty used this type of assessment in at least some of their classes. The differences among science faculty were statistically significant, $\chi^2(4, N = 2,755) = 30.128, P = .000$. See Table 2.

Grading Practices for Summative Purposes

Grading on a Curve

Faculty graded on a curve somewhat less than they used competency-based grading. About one-fourth of chemistry (27.2%) and physics faculty (25.6%) grade on a curve in “all” of their classes, compared to a smaller proportion of biology faculty (10.2%). This pattern is consistent when data is collapsed for the categories of faculty’s teaching in “all” and “some” of their classes, indicating that a greater proportion of physics and chemistry faculty grade on a curve than biology faculty (52.7% and 45.8 vs. 27.4%). The differences among science faculty disciplines were statistically significant, $\chi^2(4, N = 2,755) = 186.056, P = .000$. See Table 3.

Competency-Based Grading

About one-third of faculty from all three science disciplines used competency based grading in “all” of their classes, and just over half of faculty overall used it in “some” or “all” of their classes. There were small but statistically significant differences among science faculty disciplines, with slightly fewer chemistry faculty using this strategy than other science disciplines, $\chi^2(4, N = 2,754) = 10.044, P = .040$. See Table 3.

Conclusions and Recommendations

Educators’ ideas about assessment have changed over the last several decades, expanding the variety of possible assessment options and increasing our understanding of how these assessments can enhance teaching and learning. Each type of assessment tool is suited for a particular assessment purpose. There is a continuum of open-endedness with these assessment types that ranges from forced-choice responses of multiple-choice items, to students

supplying their own brief answer, to more elaborated responses of essays and term papers, the latter allowing students to provide the thesis as well as the organizational structure of their response. When implemented in a way that promotes feedback and formative aspects, each type of assessment provides valuable information about student learning and has intrinsic learning potential. This variety provides the necessary tools to capture and promote a wide range of science learning which is possible in an inquiry-rich science curriculum.

Recommendations for Science Education at the College Level

Recommendations are offered for college science faculty based on the assessment practices reported in this study in light of possible explanations for these findings.

Promote Implementation of a Wide Repertoire of Assessment Strategies in College Science, Particularly Among Chemistry and Physics Faculty

In the present study examining the practices of college science faculty, statistically significant differences in assessment and grading were found among the science disciplines of biology, chemistry and physics. A greater proportion of biology faculty used a wider repertoire of assessment types than physics or chemistry faculty. Biology faculty were more likely to use multiple-choice exams as well as assessments that are more open ended and provide student feedback, such as multiple drafts of written work and peer evaluation.

Reasons for explaining these differences might be related to faculty’s perceptions of the disciplines they teach. For example, the present study’s findings indicate that less than half of physics and chemistry faculty use assessments that require students to express their ideas in writing through an extended-response format such as essay answers or term papers. Since more faculty opt for short answer or multiple-choice items to assess knowledge in these disciplines, faculty might believe that the answers to the problems or questions are what are most valuable to ascertain. However, the thinking behind their students’ answers might provide important information to make a valid summative judgment about understanding as well as provide faculty formative data to improve student learning. Writing assessments could be valuable for any science discipline allowing students’ argumentation to be assessed not only during class discussions and instructional activities but also in summative evaluation. Teaching science through inquiry and being consistent with the nature of science necessitates using assessments that measure both the products and process of scientific knowing (Duschl

Table 3 Grading practices used by science faculty

	Biology		Chemistry		Physics		χ^2
	N	%	N	%	N	%	
Grading on a curve							
All	157	10.2	110	27.2	209	25.6	186.06**
Some	264	17.2	75	18.6	221	27.1	
None	1114	72.6	219	54.2	386	47.3	
Competency-based grading							
All	536	34.9	142	35.1	300	36.8	10.04*
Some	377	24.6	76	18.8	166	20.4	
None	622	40.5	186	46.0	349	42.8	

* $P < 0.05$, ** $P < 0.001$

2003) where students' scientific reasoning is demonstrated. Although assessing these learnings can be challenging, the effort is worthwhile if important aspects of learning are captured.

Helping faculty understand the usefulness of various types of assessment for their subject discipline might be a productive focus of faculty development. Because validity and reliability are more difficult to accomplish with open-ended items, it is critical to design rubrics that capture criteria aligned with learning goals and conduct reliability checks to ensure the consistency of the scoring process. By increasing awareness of the benefits of using a variety of assessment types and emphasizing ways to improve the technical quality of each item type, faculty will be better equipped to assess an assortment of learning outcomes.

Support College Science Faculty's Broader Assessment Repertoire by Promoting Faculty's Understanding of the Relationship Between Assessment and Learning

To accomplish learning purposes through assessments, strategies such as peer evaluation and students' revision of their work could be added to the assessment repertoire that faculty already utilize. Giving these strategies a place in the postsecondary curriculum can help faculty monitor students' understanding and give students a chance to build their ideas. In the present study, it appears that formative assessment tools might be underutilized for faculty of all science disciplines. For example, practices that promote student learning such as peer evaluation were reported for less than one-half of faculty of any science discipline, and the practice of students submitting multiple drafts of their work were used by less than one-third of science faculty.

There are mixed results in terms of the extent that formative assessment strategies are being used effectively in college curricula, according to the postsecondary assessment literature. Faculty might perceive time constraints, limited resources, and large class sizes of undergraduates as obstacles to utilizing such strategies. However, creative solutions to these challenges have been found by faculty such as Mazur (1997) whose approach called *Peer Instruction* is an example of a flexible approach for the postsecondary level that can be implemented in combination with other methods, even in a large section of an introductory science course.

Heady (2000) who has synthesized assessment recommendations from several standards documents encourages college science instructors to nurture as well as to evaluate student learning. However, she questions the degree that it is occurring at the postsecondary level. Yorke (2003) also supports increased efforts to implement formative assessment strategies, suggesting that acquiring content knowledge of subject disciplines has been emphasized more than

the developmental or cognitive needs of college students. A study by Tomanek et al. (2008) indicated that factors influencing teachers' choice of assessments for formative purposes included the characteristics of the task and the characteristics of the students. Tomanek et al. informs interpretation of the findings in two ways. First, the explanation for differences in terms of usage of formative assessments by biology, chemistry and physics faculty might be related to their perceptions of the specific learning tasks unique to each discipline. For example, biology faculty might view their discipline as being more varied in terms of the types thinking required and the amount of creativity allowed in demonstrating their understanding, prompting faculty to adopt a greater range of assessment types than physics and chemistry faculty. Second, the findings of Tomanek et al. indicated that the characteristics of students such as teachers' perceptions of their ability might influence assessment choices. It is possible that such perceptions of college faculty might be different than K–12 teachers because of the different thinking level faculty perceive of students. Although at every level, even when alternative assessments were implemented, the value of the assessment is sometimes not fully realized because of the manner in which it was implemented. For example, in a study of 5th graders' science notebooks, there was no evidence that teachers provided feedback or wrote comments in the notebooks (Ruiz-Primo et al. 2004) that would have enhanced the alternative assessment's learning potential. Although the assessment had promise in regards to enhancing the process skills of science students, the lack of feedback from the teacher reduced the benefits of this assessment.

Promote Grading and Reporting Practices that Align with the Purposes of Classroom Assessment

Competency-based grading is consistent with the current focus on aligning assessment with learning standards, part of the accountability movement in education. Grading on a curve is useful to compare or rank students in a classroom; however, the practice does not provide a direct measurement of student competence for a standard or learning objective. In the present study, a greater proportion of physics and chemistry than biology faculty graded on a curve in some or all of their classes (52.7% physics and 45.8% chemistry versus 27.4% biology). Although it is not clear from the data why chemistry and physics faculty were more likely to curve students' scores, a few plausible explanations might include: (a) low mean scores on physics and chemistry exams than biology exams that might prompt faculty to curve scores upward, (b) a perception of chemistry and physics faculty that not all students should receive a passing grade, prompting faculty to grade on a curve to adjust scores lower, or (c) a

traditional practice carried forward based on faculty's own prior educational experiences in their science discipline. Mazur (1997) recommends "absolute grading" so that students are more willing to share ideas and so that teachers can foster an environment conducive to collaboration rather than competition. While competency-based grading might still be underutilized in view of the emphasis of accountability reform efforts, findings from the present study indicate that about half of faculty from all science disciplines used competency-based grading at least some of the time.

While no array of assessments can perfectly measure students' understanding, using a wide variety of assessment tools takes science educators closer to that goal. Results from other studies indicate that professional development might prove effective in the areas described. For example, professional development was successful in enhancing teachers' assessment practices in a study of teachers progressing through the national board certification process (Sato et al. 2008). Teachers in the study developed a wider repertoire of assessments, including use of formative assessment strategies. Perhaps the key to increasing use of formative assessment involves understanding the reasoning behind teachers' assessment choices, as was examined for teachers at the K–12 level by Tomanek et al. (2008), a worthwhile line of inquiry that might also inform professional development practices at the college level.

Recommendation for Future Research on Assessment at the Postsecondary Level

This study provided a large-scale descriptive overview of the types of assessments college science faculty implement. NSOPF provides a nationally representative sample of science faculty and is the largest database of faculty information in existence. However, the database does not contain samples of specific assessment items used by faculty or provide information about how and when during the instructional process the assessments were implemented. In addition, there could be variation in how faculty classified their classroom assessments when interpreting the questionnaire. Some assessment types such as various types of problems might be difficult to categorize. It is recommended that future research in college science assessment focus on the specific nature and context of assessments, including the cognitive level of items, and implementation methodology that might enhance formative aspects of learning.

References

- Atkin JM, Black P (2003) Inside science education reform: a history of curricular and policy change. Teachers College Press, New York
- Bass KM, Glaser R (2004) Developing assessments to inform teaching and learning. CSE Report 628. Los Angeles, CA: University of California, Los Angeles, Graduate School of Education and Information Studies, National Center for Research on Evaluation, Standards, and Student Testing, Center for the Study of Evaluation. Also available at www.cse.ucla.edu/reports/R628.pdf
- Black P (2003) The importance of everyday assessment. In: Atkin JM, Coffey J (eds) Everyday assessment in the science classroom. NSTA Press, Washington, DC
- Black P, Wiliam D (1998) Assessment and classroom learning. *Assess Educ Princ Policy Pract* 5:7–74
- Butz WP, Bloom GA, Gross ME, Kelly TK, Kofner A, Rippen HE (2003) Is there a shortage of scientists and engineers: how would we know? Issue Paper on Science and Technology. RAND, USA
- Chappuis S, Stiggins RJ (2002) Classroom assessment for learning. *Educ Leadersh* 60:40–43
- Crane L, Winterbottom M (2008) Plants and photosynthesis: peer assessment to help students learn. *J Biol Educ* 42:150–156
- Dochy F, Segers M, Sluijsmans D (1999) The use of self-, peer and co-assessment in higher education: a review. *Stud High Educ* 24:331–350
- Duschl RA (2003) Assessment of inquiry. In: Atkin JM, Coffey J (eds) Everyday assessment in the science classroom. NSTA Press, Washington, DC
- Fellows NJ (1994) A window into thinking: Using student writing to understand conceptual change in science learning. *J Res Sci Teach* 31:985–1001
- Gronlund NE (2005) Assessment of student achievement. Allyn & Bacon, Boston
- Heady JE (2000) Assessment—a way of thinking about learning—now and in the future: the dynamic and ongoing nature of measuring and improving student learning. *J Coll Sci Teach* 29:415–421
- Hestenes D, Wells M (1992) A mechanics baseline test. *Phys Teach* 30:159–166
- Hogan TP, Murphy G (2007) Recommendations for preparing and scoring constructed-response items: what the experts say. *Appl Meas Educ* 20:427–441
- Jacobs LC (1992) Developing and using tests effectively: a guide for faculty. Jossey-Bass, San Francisco
- Jiménez-Aleixandre MP, Erduran S (2008) Argumentation in science education: an overview. In: Erduran S, Jiménez-Aleixandre MP (eds) Argumentation in science education: perspectives from classroom-based research. Springer, New York, pp 3–27
- Linn RL, Baker EL, Betebenner DW (2002) Accountability systems: implication of requirements of the no child left behind act of 2001. *Educ Res* 31:3–16
- Maclellan E (2004) How convincing is alternative assessment for use in higher education? *Assess Eval High Educ* 29:311–321
- Martinez MW (1999) Cognition and the question of test item format. *Educ Psychol* 34:207–218
- Mazur E (1997) Peer instruction: a user's manual. Prentice Hall, Upper Saddle River, NY
- McDermott LC, Redish EF (1999) Resource letter: PER-1: physics education research. *Am J Phys* 67:755–767
- National Research Council (NRC) (1996) National science education standards. National Academy Press, Washington, DC
- Palomba CA (1999) Assessment essentials: planning, implementing, and improving assessment in higher education. Jossey-Bass, San Francisco
- Popham J (2003) The seductive allure of data. *Educ Leadersh* 60:48–51
- Richmond G, Parker J, Urban-Luraine M, Merritt B, Merritt J, Patterson R (2008) Assessment-informed instructional design to support principled reasoning in college-level biology. Paper

- presented at the NARST Annual International Conference, Baltimore, MD
- Roby JF, White BJ (1990) Laboratory practical exams in the biochemistry lab course. *J Chem Educ* 67:600–601
- Ruiz-Primo MA, Furtak EM (2007) Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *J Res Sci Teach* 44:57–84
- Ruiz-Primo MA, Li M, Ayala C, Park R, Shavelson RJ (2004) Evaluating students' science notebooks as an assessment tool. *Int J Sci Educ* 26:1477–1506
- Sadler DR (2005) Interpretations of criteria-based assessment and grading in higher education. *Assess Eval High Educ* 30:175–194
- Sato M, Wei RC, Darling-Hammond L (2008) Improving teachers' assessment practices through professional development: the case of National Board Certification. *Am Educ Res J* 45:669–700
- Shepard LA (2000) The role of assessment in a learning culture. *Educ Res* 29:4–14
- Slater TF (1997) The effectiveness of portfolio assessments in science: integrating an alternative, holistic approach to learning into the classroom. *J Coll Sci Teach* 26:315–318
- Stiggins RJ (1991) Facing the challenges of a new era of educational assessment. *Appl Meas Educ* 4:263–273
- Takao AY, Kelly GJ (2003) Assessment of evidence in university students' scientific writing. *Sci Educ* 12:341–363
- Tomanek D, Talanquer V, Novodvorsky I (2008) What do science teachers consider when selecting formative assessment tasks? *J Res Sci Teach* 45:1113–1130
- US Department of Education (2001) No Child Left Behind Act of 2001. Report 107–334. House of representatives, 107th Congress, 1st Session
- US Department of Education, National Center for Education Statistics (2002) National Study of Postsecondary Faculty (NSOPF:99) Methodology report. NCEES 2002-154 by S. Y. Abraham, D. M. Steiger, M. Montgomery, B. D. Kuhr, R. Tourangeau, B. Montgomery, and M. Chattopadhyay. Project Officer: L. J. Zimble. Washington, DC:NCEES
- Wiggins G (1998) *Educative assessment: designing assessment to inform and improve student performance*. Jossey-Bass, San Francisco
- Yager RE (1991) The constructivist learning model: towards real reform in science education. *Sci Teach* 58:52–57
- Yorke M (2003) Formative assessment in higher education: moves towards theory and the enhancement of pedagogic practice. *High Educ* 45:477–501