

Development of a Certification Exam to Assess Undergraduate Students' Proficiency in Biochemistry and Molecular Biology Core Concepts

Quira Zeidan,¹ Jennifer Loertscher,² Adele J. Wolfson,³ John T. Tansey,⁴ Erika G. Offerdahl,⁵ Peter J. Kennelly,⁶ Daniel R. Dries,⁷ Victoria Del Gaizo Moore,⁸ Diane M. Dean,⁹ L. Michael Carastro,¹⁰ Sachel M. Villafaña,¹¹ and Ludmila Tyler^{12*}

¹American Society for Biochemistry and Molecular Biology, Rockville, MD 20852; ²Department of Chemistry, Seattle University, Seattle, WA 98122; ³Department of Chemistry, Wellesley College, Wellesley MA 02481; ⁴Chemistry Department, Program in Biochemistry and Molecular Biology, Otterbein University, Westerville, OH 43081; ⁵School of Molecular Biosciences, Washington State University, Pullman, WA 99164; ⁶Department of Biochemistry, Virginia Polytechnic Institute & State University, Blacksburg, VA 24061; ⁷Department of Chemistry and Biochemistry, Juniata College, Huntingdon, PA 16652; ⁸Department of Chemistry, Elon University, Elon, NC 27244; ⁹Department of Chemistry, University of Saint Joseph, West Hartford, CT 06117; ¹⁰Department of Chemistry, Biochemistry & Physics, University of Tampa, Tampa, FL 33606; ¹¹Department of Chemistry and Biochemistry, California State University Fullerton, Fullerton, CA 92831; ¹²Biochemistry and Molecular Biology Department, University of Massachusetts Amherst, Amherst, MA 01003

ABSTRACT

With support from the American Society for Biochemistry and Molecular Biology (ASBMB), a community of biochemistry and molecular biology (BMB) scientist-educators has developed and administered an assessment instrument designed to evaluate student competence across four core concept and skill areas fundamental to BMB. The four areas encompass energy and metabolism; information storage and transfer; macromolecular structure, function, and assembly; and skills including analytical and quantitative reasoning. First offered in 2014, the exam has now been administered to nearly 4000 students in ASBMB-accredited programs at more than 70 colleges and universities. Here, we describe the development and continued maturation of the exam program, including the organic role of faculty volunteers as drivers and stewards of all facets: content and format selection, question development, and scoring.

Several national initiatives for improving the education of undergraduate science, technology, engineering, and mathematics (STEM) majors explicitly call for attending not only to *how* students are taught, but also to the *role of assessment* in the preparation of the next generation of scientists (American Association for the Advancement of Science [AAAS], 2011; President's Council of Advisors on Science and Technology, 2012). Assessment is critical for diagnosing and scaffolding student learning during instruction. Programmatically, coordinated assessment efforts enable departments to measure student learning and evaluate the efficacy of instructional practices and curricular improvements (Middaugh, 2010). Professional societies, which have traditionally promoted the development of scientists' research careers, have a potentially significant role to play in supporting undergraduate STEM learning through improved assessment (Hutchings, 2011) by describing best practices in society publications, providing professional development and resources, and developing instruments to assess learning in the discipline. In this *Essay*, we report on the continuing efforts of one professional society, the American Society for Biochemistry and Molecular Biology

Erin L. Dolan, *Monitoring Editor*

Submitted Dec 5, 2019; Revised Feb 8, 2021; Accepted Feb 11, 2021

CBE Life Sci Educ June 1, 2021 20:es6

DOI:10.1187/cbe.19-12-0265

Conflict of interest statement: The authors of this work have been and continue to be involved in the development of the ASBMB certification exit exam. The ASBMB charges a small fee for the certification exam to partly defray the costs of printing and shipping.

*Address correspondence to: Ludmila Tyler (ltyler@biochem.umass.edu).

© 2021 Q. Zeidan et al. CBE—Life Sciences Education © 2021 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

"ASCB®" and "The American Society for Cell Biology®" are registered trademarks of The American Society for Cell Biology.

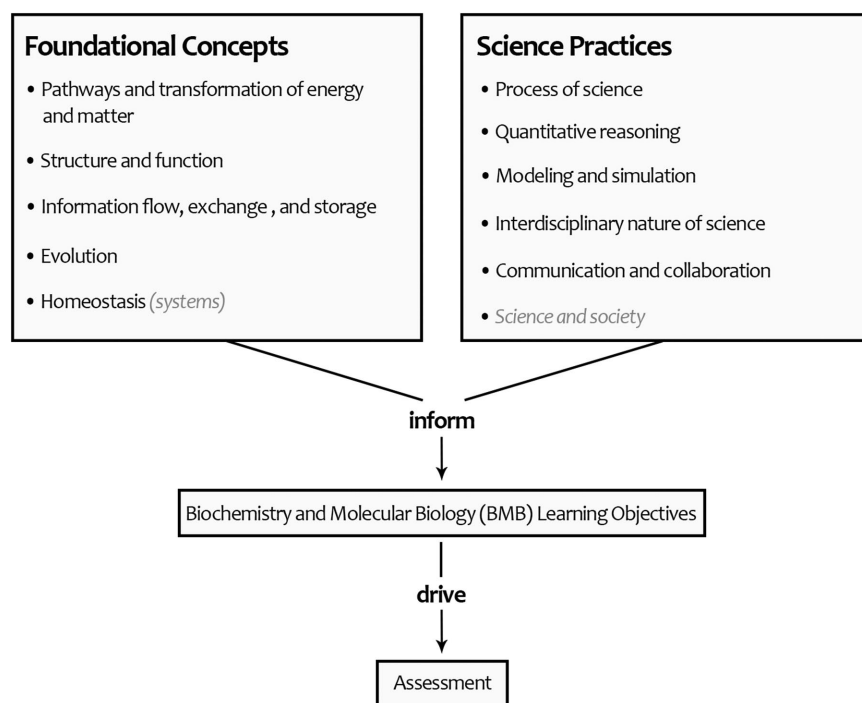


FIGURE 1. Foundational concepts and scientific practices inform the learning objectives that drive the development of the ASBMB certification exam as an assessment tool for measuring undergraduates' proficiency in BMB. The uppermost boxes illuminate how the foundational concepts and scientific practices identified by the ASBMB community map onto the equivalents articulated in the Vision and Change initiative. With the exception of those in gray, italicized font, the concepts and practices listed are emphasized in both efforts.

(ASBMB), to develop and implement a discipline-based certification exam for undergraduate biochemistry and molecular biology (BMB) majors. Specifically, we provide a descriptive account that focuses on the exam process: the grassroots origins of the ASBMB certification exam, the iterative approach through which evidence of validity continues to be collected, and the implications and future directions of such an effort by a professional society for undergraduate STEM education. We opted to publish this description as an *Essay* instead of an article, because our aim is to highlight the community-driven nature of this approach to assessment development and testing, rather than to provide a more traditional report of the development of an assessment tool.

ORIGINS OF THE ASBMB CERTIFICATION EXAM

In 2011, the AAAS publication *Vision and Change* articulated core concepts for biological literacy and core competencies of disciplinary practice in the life sciences (AAAS, 2011). Around the same time, members of the BMB education community collaboratively identified foundational concepts and skills specific to BMB as a discipline (Tansey *et al.*, 2013; White *et al.*, 2013; Wright *et al.*, 2013). The concepts and skills identified by AAAS and the BMB community exhibit substantial overlap (Figure 1). Brownell *et al.* (2014) subsequently outlined how the core concepts of *Vision and Change* could be interpreted for general biology courses, and several professional societies have similarly

interpreted the concepts for their own sub-disciplines (American Society of Plant Biologists, 2012; Merkel, 2012). After refining the inventory of core BMB concepts and skills and articulating a set of aligned learning objectives (Tansey *et al.* 2013; White *et al.*, 2013), ASBMB applied this framework to the development of an accreditation process for undergraduate programs and a certification exam for their students. The certification exam, which we describe here, is designed to assess proficiency in core concepts and skills as students near completion of a biochemistry and/or molecular biology major. Other prominent examples of professional societies providing criteria for accreditation and access to curricular and assessment resources include the Accreditation Board for Engineering and Technology (www.abet.org), the Accreditation Council for Education in Nutrition and Dietetics (www.eatrightpro.org/acend) of the Academy of Nutrition and Dietetics, the Accreditation Commission for Education in Nursing (www.acenursing.org), and the American Chemical Society (ACS, www.acs.org/content/acs/en.html).

Historically, several of the assessment tools that are widely used in BMB programs have come from ACS. Through its Division of Chemical Education, ACS has long assisted programs in collecting and analyzing data via their affiliated Examinations

Institute, which first offered a true–false general chemistry national exam in 1934 (Emenike *et al.*, 2013; Brandriet *et al.*, 2015). Since then, ACS has substantially expanded its spectrum of examinations to encompass chemistry-related topics ranging from analytical chemistry to chemical health and safety and, as of 2007, biochemistry (<https://uwm.edu/acs-exams>). Modern ACS exams, which employ multiple-choice items designed to target a variety of cognitive levels (Brandriet *et al.*, 2015), have been extensively analyzed for both item performance (Schroeder *et al.*, 2012) and item format (Brandriet *et al.*, 2015).

Other available assessment tools include concept inventories, research-based assessments for formatively informing instructional design and monitoring student progress across a series of courses within a curriculum. Multiple-choice concept inventories have been developed to probe students' understanding related to the molecular life sciences (Howitt *et al.*, 2008), foundational concepts in biochemistry (Villafañe *et al.*, 2011; Xu *et al.*, 2017), enzyme–substrate interactions (Bretz and Linenberger, 2012), genetics (Smith *et al.*, 2008), and molecular and cell biology (Shi *et al.*, 2010). A small number of constructed-response assessments are also available (Villafañe *et al.*, 2016). The Biology Card Sorting Task (Smith *et al.*, 2013) assesses the degree to which students' conceptual knowledge in biology is organized in expert-like structures and has been suggested as a measure of students' conceptual development over time. The General Biology–Measuring

Achievement and Progression in Science (GenBio-MAPS) assessment evaluates student understanding of core concepts at critical junctures in undergraduate biology programs (Couch *et al.*, 2019). Notably, while some of the tools assess central BMB concepts, most focus on introductory-level content and target only one aspect of BMB. Thus, despite their strengths, none of these assessment tools is entirely suited for measuring the conceptual understanding and competencies spanning a BMB program.

This latter point is important, because modern biochemistry and molecular biology have coalesced into a distinct discipline well beyond the simple intersection of chemistry and biology. One cannot fully understand the form and function of a biological molecule or system without considering biological context; chemical properties, structure, and reactivity of components; and evolutionary history. That is, the kinetic parameters and the pattern of expression are both important facets of an enzyme. While one is “chemical” and the other “biological,” integrating the two presents a far richer picture of the enzyme than either perspective can alone. The ACS biochemistry exam focuses heavily on more “chemical” topics such as energetics and metabolism and macromolecular structure–function, and less on topics of information transfer and molecular evolution that constitute equally vital components of BMB curricula. An exam addressing the full spectrum of BMB must emphasize both perspectives and their interrelationship within a living organism.

There is, furthermore, growing emphasis on instruction and assessments that move beyond traditional insular approaches to support students in understanding crosscutting concepts such as those inherent to BMB (Lavery *et al.*, 2016; Bain *et al.*, 2020). The ASBMB certification exam, which is available annually to ASBMB-accredited BMB programs and their students, addresses competencies as well as factual knowledge. Exams are constructed on an annual basis by teams of experts from a bank of questions that have been subjected to an iterative design process intended to produce items that target one of four core concept and skill areas (energy and metabolism, structure–function relationships, information storage and transfer, and analytical/quantitative reasoning skills; www.asbmb.org/education/core-concept-teaching-strategies/foundational-concepts) at a defined level of cognitive processing.

This *Essay* describes how the BMB community has coalesced to develop, refine, and ultimately sustain an assessment tool tailored for the discipline. In addition to outlining the community-driven process by which the ASBMB certification exam is constructed, administered, and scored, we seek to highlight ways in which principles of assessment instrument design are being used to elevate the quality of the exam, in alignment with best practices articulated by the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (AERA, APA, and NCME) and others (AERA *et al.*, 2014; Bandalos, 2018). We present an evolving body of evidence to support the validity of items in the instrument. Because a distinct exam is constructed each year, we describe the 2019 exam in detail, including an analysis of item difficulty and discrimination, as a concrete example for readers. Finally, we discuss the implications and future of the ASBMB certification exam.

CONTEXT AND PURPOSE OF THE EXAM: THE ASBMB ACCREDITATION PROGRAM

In 2013, ASBMB began offering accreditation for undergraduate programs in BMB and the related molecular life sciences whose features and infrastructure fulfill the basic expectations of the society (Dean *et al.*, 2018; Del Gaizo Moore *et al.*, 2018). One of the foundational objectives of the accreditation program was the establishment of an independent, outcomes-based credential by which the society could recognize students who exhibit a solid foundation in BMB. Such a credential would enable students to certify their proficiency according to an external standard, independent of their colleges’ or universities’ reputations. Further, it was recognized that the independently generated data yielded by the certification exam could serve as a valuable resource for programmatic assessment.

IDENTIFICATION OF FOUNDATIONAL CONCEPTS IN BMB

A critical first step in instrument development is clear articulation of learning targets to be assessed. To this end, BMB scientist-educators were invited to a series of two dozen small workshops held across the United States from 2010 to 2014. These workshops, which were funded by a Research Coordination Networks for Undergraduate Biology Education (RCN-UBE) grant from the National Science Foundation (award no. 0957205), provided opportunities for several hundred scientist-educators to define an inventory of BMB core concepts likely to be valued across the BMB community. A consensus coalesced around four core concept and skill areas: energy and metabolism, information storage and transfer, macromolecular structure and function, and use of scientific practices including quantitative analysis and analytical reasoning (Mattos *et al.*, 2013; Tansey *et al.*, 2013). In addition, the community explicitly recognized that these four areas are permeated and linked by the underlying principles of evolution and homeostasis. This consensus among disciplinary experts for the areas targeted by the exam provides evidence of content validity for the assessment. Today, these four core concept and skill areas continue to define the domain of the certification exam and form the foundation for question development (Figure 1).

BROAD COMMUNITY ENGAGEMENT IN EXAM DEVELOPMENT AND SCORING

The involvement of a large community of BMB scientist-educators has been essential in all aspects of exam development, administration, and scoring. The initial cohort consisted of a small, eight-member group that, with the support of a grant from the Teagle Foundation, was trained by external experts in assessment techniques during a series of three weekend-long workshops. As the program has grown (Figure 2), additional volunteers have been recruited: at workshops and conferences, via articles in the society’s news magazine, and through email invitations to both individual ASBMB members and directors of accredited programs.

Question-writing teams in each of the BMB core areas were established early in the exam development process. Attendees at some of the later RCN-UBE workshops (described earlier) also generated questions, and many of these BMB scientist-educators subsequently joined ASBMB’s question-writing and exam-scoring teams. More recently, dedicated question development

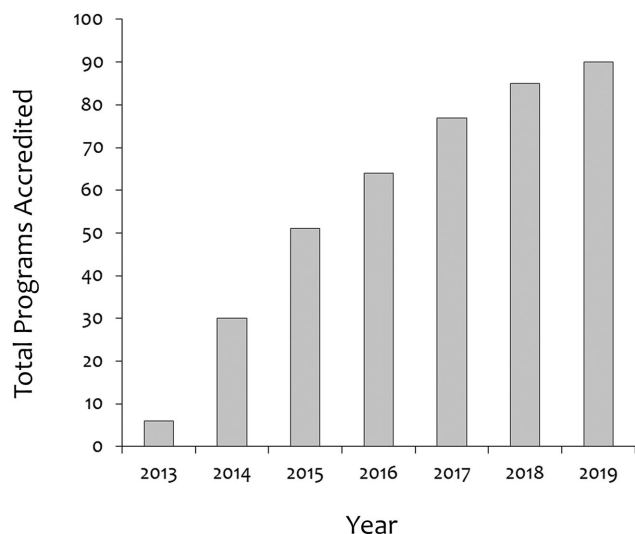


FIGURE 2. The number of ASBMB-accredited undergraduate programs has grown each year. Shown is the number of programs accredited at the ends of the indicated calendar years since the ASBMB's Accreditation Program was launched in 2013. Programs accredited in the Fall are eligible to participate in the certification exam in the Spring of the following year.

workshops have become a regular feature of both the society's annual meetings and its biennial small education conferences.

To date, approximately 120 individuals have been involved in question development and/or scoring, many of whom have volunteered over multiple years (Supplemental Material 1). The core cadre of faculty volunteers has been supplemented by a few graduate students and postdoctoral scientists involved in undergraduate BMB education. The professional affiliations of these volunteers range from small, primarily undergraduate institutions to large research universities (Supplemental Material 1). Cultivating a community of volunteers from a variety of institutions brings a range of expert perspectives to the creation and review of exam questions, with the added benefits of distributing the workload and increasing national engagement with the certification exam.

The large volunteer community also constitutes a vital source of validity evidence used to determine the degree to which data support the interpretation of exam scores (AERA *et al.*, 2014; Reeves and Marbach-Ad, 2016). As described later, continually collecting expert feedback from question-writing and exam-scoring teams throughout the exam development process provides validity evidence based on test content. Experts are further involved in evaluating validity evidence based on response processes, specifically in using student responses on pilot questions to inform revisions. Through continuous, organic input, the volunteer community elevates the quality of the exam over time. In recognition of their contributions, the society has designated these BMB scientist-educators ASBMB Education Fellows.

CRITERIA FOR QUESTION DEVELOPMENT

Since 2013, ASBMB's exam development community has engaged in an iterative process to develop a bank of questions and corresponding rubrics targeting the BMB core concept and skill areas at lower and higher levels of cognitive processing. Starting with well-defined learning objectives, question development teams create questions and rubrics that assess a single learning objective within their assigned concept or skill area. These questions require a specifically delineated response, described by an accompanying rubric. Examples of both appropriately targeted and unacceptably vague objectives for developing exam questions are shown in Table 1. To probe different degrees of cognitive processing, development teams apply Bloom's taxonomy (Bloom, 1956; Crowe *et al.*, 2008). Because the taxonomy is not necessarily hierarchical past the third of the six classification levels (Crowe *et al.*, 2008), the teams use it to distinguish between questions that require only minimal cognitive processing (i.e., lower-order cognitive skills, or LOCS) versus more substantial cognitive processing (i.e., higher-order cognitive skills, or HOCS). LOCS questions most often assess knowledge recall or the ability to demonstrate basic comprehension of biochemical concepts, for example, recognition of a correct answer in an array of alternatives. An example LOCS question testing a student's ability to recognize the correct answer is shown in Figure 3. The corresponding rubric (Figure 3) is simple, and student responses can be scored quickly. In

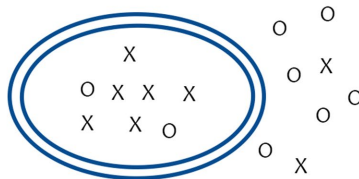
TABLE 1. Examples of acceptable and unacceptable question frameworks

	Concept area	Unacceptable	Acceptable
Example 1	Energy and metabolism	Does a student understand thermodynamic coupling?	Given a list of chemical reactions and their delta G values, can a student select an appropriate reaction to couple to a given, thermodynamically unfavorable one?
Example 2	Macromolecular structure, function, and assembly	Does a student understand how biological molecules form three-dimensional structures?	Given a list of examples of folding of biological molecules and assembly of macromolecular structures, can a student identify those examples in which the maximization of entropy is the predominant thermodynamic driving force?
Example 3	Information storage and transfer	Does a student understand the central dogma of DNA being transcribed to RNA and mRNA being translated into protein?	Can a student recognize a frameshift mutation and explain its impact on protein function?
Example 4	Scientific method, including quantitative reasoning	Does a student understand the concept of pH?	Can a student calculate the pH of a sufficiently described buffer system?

Example of a LOCS Question

Circle any of the transport events listed below which is/are thermodynamically favorable under the conditions illustrated to the right.

- Compound X being transported into the cell.
- Compound O being transported into the cell.
- Compound X being transported out of the cell.
- Compound O being transported out of the cell.



Example of a LOCS Rubric

Highly Proficient Answer:

Student circles choices **ii** and **iii** only.

Proficient Answers Include:

Student circles choice **ii** only.

OR

Student circles choice **iii** only.

Not Yet Proficient Answers Include:

All other combinations.

FIGURE 3. Example of a LOCS exam question and rubric. A Concept/Skill Area 4 LOCS question in a multiple-select format is shown above the corresponding rubric. The diagram that is part of the question depicts an oval-shaped cell bilayer membrane, as well as two compounds (X and O).

contrast, HOCS questions probe conceptual understanding by requiring application of knowledge to novel contexts, evaluation of information, and synthesis of a quantitative/qualitative solution or explanation, for example, design and explanation of an experimental approach. An example HOCS question is shown in Figure 4. This question requires that a student interpret the data presented and formulate an acceptable explanation. The rubric (Figure 4) is more complex, and raters must carefully assess the depth of understanding conveyed in a student's response. Notably, Bloom's taxonomy should not be conflated with item difficulty (Crowe *et al.*, 2008; Lemons and Lemons, 2013; Arneson and Offerdahl, 2018). Rather, Bloom's taxonomy serves as a guide to construct questions that evaluate knowledge of foundational concepts and disciplinary skills (e.g., data analysis and interpretation) across levels of cognitive processing.

QUESTION REFINEMENT AND COLLECTION OF VALIDITY EVIDENCE

Each year, drafts of prospective questions undergo iterative cycles of review and refinement by teams of question developers (Figure 5). These teams first determine whether the questions are correct, clear, concise, and focused on targeted learn-

ing objectives. The teams also evaluate whether questions may be improved by the inclusion of figures, diagrams, or tables. In 2017, a question-writing guide (Supplemental Material 2) was compiled to consolidate lessons learned as a means for elevating quality and promoting uniformity across the question development process. Emphasizing principles of backward design (Wiggins and McTighe, 2005), the guide provides detailed instructions on writing clear, focused questions that are intentionally designed to elicit responses related to specific learning objectives. This document, which continues to be revised, is provided to every volunteer involved in exam development.

Once draft questions have been scrutinized for clarity and relevance, content validity evidence is further collected through a process of expert review conducted independently of the question developers, generally by members of the scoring teams. The fresh and varied perspectives of the scoring teams have proven to be a powerful aid in identifying and removing implicit content, resolving ambiguities, simplifying phrasing, and highlighting instances where an illustrative figure would be useful.

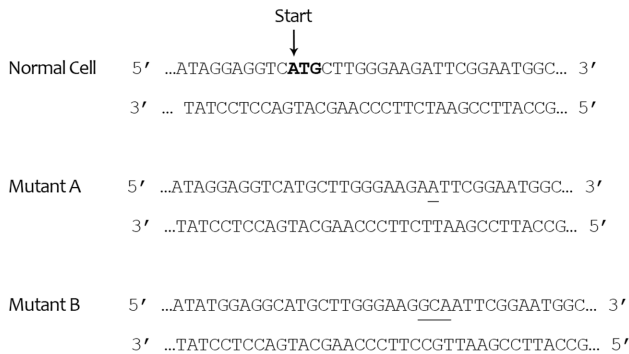
Next, students' written responses to the piloted questions are collected and analyzed. This information provides insight into how students are processing the question and is used to generate suggestions for improvements. The approach of examining student answers to pilot questions is a method for collecting validity evidence of the response process, because it provides "records that monitor the development of a response" (Padilla and Benítez, 2014, p. 139). The original and revised questions are then submitted to the exam steering committee for discussion and, if approved, are deposited in the exam question bank. Alternatively, piloting of the revised version may be prescribed. Our iterative question evolution process is summarized in Figure 5 and illustrated by the example described in Supplemental Material 3.

ANNUAL EXAM CONSTRUCTION

Each year, construction of the ASBMB certification exam is overseen by an exam steering committee consisting of BMB scientist-educators possessing multiple years of experience with the exam. Typically, 12 questions are chosen for inclusion in each administration of the exam. These questions are distributed approximately equally across the four core concept and skill areas (Table 2), using one LOCS question and one or two HOCS questions to assess each area (Bloom, 1956; Zoller, 1993; Crowe *et al.*, 2008). Annually, one of the concept areas is represented by two, instead of three, questions, to allow time for a pilot question within the 60-minute exam period.

Example of a HOCS Question

Cells in culture were exposed to UV radiation for the purpose of generating DNA mutations. Afterwards, two cell lines were isolated, each with a different insertion mutation in the coding region of the gene for protein X. Shown below are the sequences for the corresponding regions of DNA for a normal (non-irradiated control) cell and the two mutant cells, A and B. The translation **start codon** for protein X is highlighted in bold type, and the newly inserted base pairs in mutants A and B are underlined.



Explain why the single insertion in mutant A is more likely to result in a dysfunctional protein product than the nucleotide triplet that was inserted in mutant B.

Example of a HOCS Rubric

Highly Proficient Answers Include:

Mutation A produces a frameshift that alters the codons for every subsequent amino acid in the peptide. **AND** Mutant B simply inserts an additional amino acid into the polypeptide chain while keeping the remainder of the protein's sequence intact. **AND** The answer provides some explanation that an insertion of a single amino acid is much less likely to disrupt the structure and function of a protein than a change in multiple amino acids.

Proficient Answers Include:

Correctly identifies mutant A as a frameshift and mutant B as adding only one amino acid but then describes the change to the protein as caused by a clearly erroneous reason:
Mutant A causes a mutation at a structurally critical residue. **OR** Mutant B introduces a premature STOP codon. **OR** Mutant A must change the amino acid to a proline or glycine. **OR** Mutant B causes a frameshift mutation (or words to that effect).

Mutant A causes a frameshift mutation (or words to that effect) that introduces (or could introduce) a premature stop codon.

Mutant A alters multiple (many) amino acid residues, while mutant B alters the protein by inserting only a single amino acid.

Not Yet Proficient Answers Include:

Answer indicates that Mutant A causes a frameshift (without mentioning mutant B), but also includes one or more of the following or any other clearly erroneous statement:
Mutant A causes a mutation at a structurally critical residue. **OR** Mutant B introduces a premature STOP codon. **OR** Mutant A must change the amino acid to a proline or glycine.

Answer does not mention that mutant A causes a frameshift (or words to that effect).

FIGURE 4. Example of a HOCS exam question and rubric. A Concept/Skill Area 3 HOCS question requiring a constructed response is shown above the corresponding rubric. The diagram that is part of the question depicts three (Normal, Mutant A, and Mutant B) duplex DNA sequences.

Question formats are balanced between open-ended questions (e.g., constructed responses or mathematical solutions) and quick-scoring multiple-select and multiple-choice questions. The initial draft of the exam, with questions and rubrics, is reviewed by additional experienced volunteers, who provide feedback regarding overall exam composition, as well as individual questions and rubrics. Next, scoring volunteers review, discuss, and further polish the questions and corresponding answer keys, to ensure that the final, official version of each question is of the highest possible quality. At least one round of this refinement process occurs before a final version of the exam is approved (Figure 6).

EXAM ADMINISTRATION

After construction and final review, the exam is provided to those ASBMB-accredited programs that elect to participate. Selected practice questions with corresponding answer keys are provided to assist students in preparing for the exam (www.asbmb.org/education/certification-exam). It is left to the judgment of the individual programs to determine whether, in the context of their curricula, students are best prepared to take the exam as seniors or juniors. To date, the certification exam has typically been available during a 2-week window in the spring of each year. Programs are asked to have all eligible students take the exam during the same 60-minute period unless an accommodation is requested. Conventional proctoring practices are required, as detailed in a letter mailed to the exam administrator (Supplemental Material 4). Completed exams are then returned to ASBMB for scoring.

PROCEDURE FOR AND RELIABILITY OF EXAM SCORING

Student answers are assessed against a rubric using a three-tiered scale: 3 = highly proficient, 2 = proficient, and 1 = not yet proficient, with a score of zero given to unanswered questions. Each student response is scored by a team consisting of at least three volunteer BMB scientist-educators, who are assigned to questions based on their areas of expertise. Initially, each rater individually evaluates the answer according to the key. The scoring team then engages in collective discussion as needed. These scoring teams serve as the functional units for training of raters, collecting input for question and answer

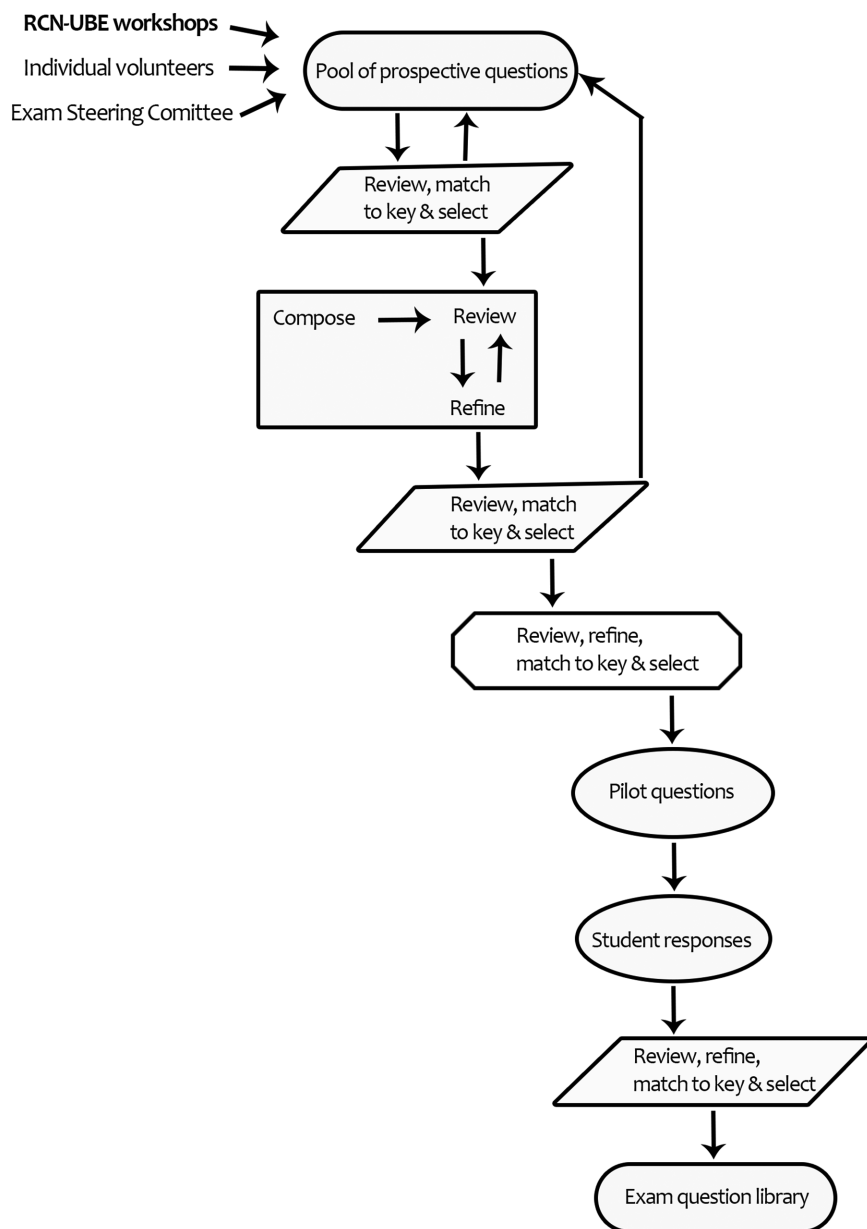


FIGURE 5. Question development process. The flowchart summarizes the iterative process by which ASBMB collects, reviews, refines, and pilots questions and their associated answer keys for use in future certification exams. Prospective questions derive from a number of sources: individuals, participants in question-writing workshops, and participants in the ASBMB-sponsored, NSF-funded RCN-UBE workshop series held from 2011 to 2016. For each step in the question development process, the group responsible for overseeing its successful completion is indicated by the geometry of the shape that encloses that step. Steps enclosed within parallelograms are overseen by the exam steering committee. The step enclosed by a rectangle, which includes public question-writing workshops, falls under the purview of the question-writing subcommittee. The step enclosed by an octagon is conducted by scoring teams. Specific work products are enclosed by rounded shapes.

key development, and evaluating student answers. Each response is assigned an overall proficiency level based on the average of the scores given by the raters (0.00–1.50 = not yet proficient, 1.51–2.50 = proficient, and 2.51–3.00 = highly profi-

cient). For instance, if one rater gave a score of “1” and two raters gave a score of “2,” the overall proficiency level of the response, 1.67, would be proficient according to these cutoffs.

The prior participation of raters in the review of questions and rubrics generally results in a robust consensus. To ensure that reasonable agreement has emerged in practice before scoring the entire question set, raters are first asked to score a subset of ~50 student responses to their assigned questions (Figure 6). These scores are used to calculate a preliminary interrater reliability – a measure of consistency among the members of the scoring team – using Fleiss’ kappa (κ ; Fleiss, 1971). The kappa statistical function ranges between 0 (perfectly opposite scores, no agreement) and 1 (complete agreement among scores). Should the preliminary κ value fall below 0.5, one or more exam team leaders will assist the raters to identify and resolve points of inconsistency, such as a failure to anticipate a particular student response, and, if necessary, further refine the rubric. The full set of exams is then scored using the final, agreed-upon rubric (Figure 6).

Because performance on the exam is intended to reflect competency across BMB, the proportion of a student’s responses evaluated as proficient or highly proficient is used to determine certification. To earn this honor, students must correctly answer (at proficient or above) a majority of the questions in at least three of the four BMB concept and skill areas or one or more questions in all four areas. The exam steering committee reviews the scores to confirm or adjust, as appropriate in a given year, the performance thresholds. Historically, a student has been expected to achieve scores of proficient or highly proficient on approximately 65% of the HOCS and 75% of the LOCS questions on the exam to qualify for certification; this threshold correlates with a score of proficient or above on ~70% of total exam questions. Certification with distinction has been awarded to students earning scores of proficient or highly proficient on approximately 83% of the exam questions. On average, approximately 42% of students have earned certification, and 13% of the total have earned certification with distinction each year (Table 3).

DESCRIPTION AND ANALYSIS OF THE 2019 EXAM

The 2019 exam was constructed with the benefit of 5 years of prior experience in exam development and scoring nearly 3000

TABLE 2. Selection of exam questions^a

Year	Energy and metabolism	Information storage and transfer	Macromolecular structure, function, and assembly	Analytical and quantitative reasoning	Pilot question ^b	Total no. of scored exam questions (pilot not included)
2014	LOCS = 2 HOCS = 1	LOCS = 0 HOCS = 4	LOCS = 3 HOCS = 1	LOCS = 0 HOCS = 2	0	13
2015	LOCS = 1 HOCS = 2	LOCS = 0 HOCS = 3	LOCS = 2 HOCS = 3	LOCS = 1 HOCS = 1	1	13
2016	LOCS = 1 HOCS = 2	LOCS = 1 HOCS = 2	LOCS = 1 HOCS = 1	LOCS = 1 HOCS = 2	1	12
2017	LOCS = 1 HOCS = 2	LOCS = 1 HOCS = 2	LOCS = 1 HOCS = 1	LOCS = 1 HOCS = 2	1	11
2018	LOCS = 1 HOCS = 1	LOCS = 1 HOCS = 2	LOCS = 1 HOCS = 2	LOCS = 1 HOCS = 2	1	11
2019	LOCS = 1 HOCS = 2	LOCS = 1 HOCS = 1	LOCS = 1 HOCS = 2	LOCS = 1 HOCS = 2	1	11

^aThe number of exam questions and the balance of questions at lower and upper Bloom's cognitive skill levels (LOCS/HOCS) in each of the four content areas have varied.

^bPilot questions can come from any of the four core concept areas at either the LOCS or HOCS level. Multiple questions may be piloted in any given year, but only one pilot question is included on any given exam.

total student responses. Thus, the 2019 exam was the result of a relatively mature process representative of refined criteria we have established for the annual ASBMB certification exam. This exam consisted of 12 questions, 11 that contributed to students' overall score plus one pilot question. As is typical, one LOCS and two HOCS questions were included for each BMB area, with the exception of information storage and transfer, for which one LOCS and only one HOCS question were included, in order to accommodate the pilot question. Six of the 11 questions required constructed responses; the remaining five had a quick-scoring multiple-select format. Table 4 summarizes the order and type of questions on the 2019 exam.

In 2019, there were 993 exams from 73 institutions scored by 53 volunteer raters. As described earlier, questions were scored by teams of three raters. Given the large number of exams in 2019, two teams were assigned to each constructed-response question, with each team scoring half of the responses. A single three-rater team scored all responses for each multiple-select question. For the purpose of this analysis, exams with missing or incomplete responses were removed, and an item analysis was performed on the remaining data set of complete exams for 2019 ($N = 904$).

Item difficulty, or the mean score, was calculated for each question. While the possible item difficulty ranged from 1.00 (most difficult) to 3.00 (least difficult), the averages on the 2019 exam ranged from 1.64 to 2.51 (Table 4). With the exception of question 9, whose average fell on the low end of the highly proficient range, the average difficulty of all other items fell within the proficient range (Table 4). These values suggest the exam questions were moderately difficult and challenged students consistently across the four concept/skill areas as intended. Developing an exam with average question scores in the proficient range is the result of a years-long process of question refinement aimed at aligning the assessment instrument with the competencies targeted for measurement.

Item discrimination analysis measures how well an item differentiates between students who score high or low on the overall exam. This analysis, which divides students into groups of high and low achievers, was calculated using the item-to-total

correlation in SPSS (Statistical Package for the Social Sciences, MAC OS v. 26.0; Kline, 2005; IBM, 2019). Table 4 shows that questions on the 2019 exam exhibit fair to excellent ability to distinguish between low- and high-achieving students (Kline, 2005).

As in previous years, the 2019 thresholds were based on the number of HOCS and LOCS questions answered correctly (at a level of proficient or highly proficient). Of the 993 students in ASBMB-accredited programs who took the exam nationwide in 2019, 412 (41.5%) achieved certification. In addition, 114 (11.5% of the total) achieved certification with distinction. These values are consistent with average percentages for student performance from 2014 to 2018 (Table 3).

EVOLUTION OF THE EXAM BASED ON DATA ANALYSIS

The construction of a new exam each year provides the opportunity for ongoing improvement as additional data are collected and analyzed. For instance, in 2019, students earned certification if they answered either five HOCS questions and three LOCS questions or six HOCS questions and two LOCS questions at proficient or above. However, subsequent item difficulty analysis revealed that some of the most difficult questions were in the LOCS category (questions 1 and 8), whereas some of the least difficult fell into the HOCS category (questions 2, 6, 9, and 11). While all ASBMB-accredited programs would be expected to support students in attaining broad proficiency across the four core concept and skill areas, other factors such as the emphasis placed on specific learning objectives in a particular curriculum may be a stronger determinant of a question's difficulty for an individual student than the nature of the question as HOCS or LOCS. Indeed, Lemons and Lemons (2013) explicitly describe difficulty and Bloom's level as distinct dimensions of a question. Thus, considering HOCS and LOCS categories separately when setting certification thresholds for the ASBMB exam may be unnecessarily complex. Analysis of item difficulty and discrimination of future exams could clarify whether or not our current system should be replaced by certification based simply on the total number of questions (at least eight of 11, or 73%) scored proficient or better.

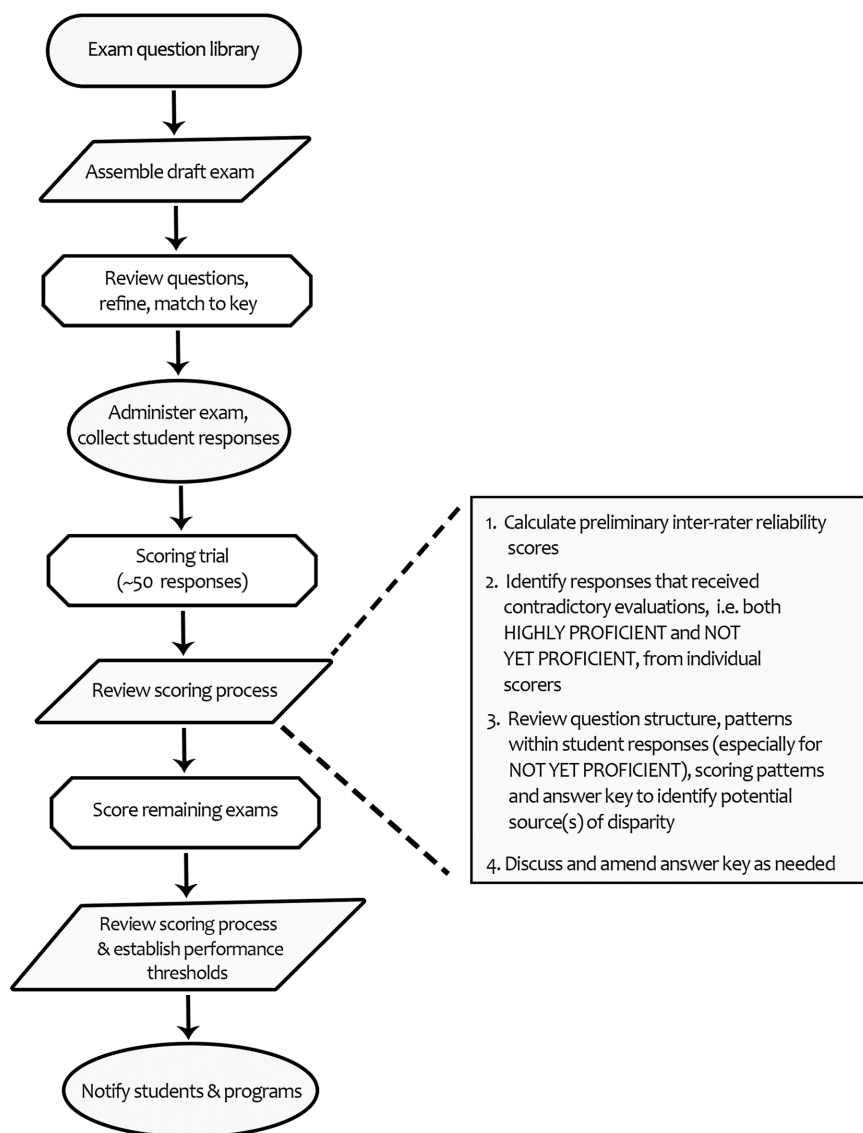


FIGURE 6. Exam construction and scoring process. The flowchart summarizes the overlapping and iterative process by which ASBMB constructs, reviews, administers, and scores its annual certification exam. The inset provides details of the review of the scoring process. For each step in the question development process, the group responsible for overseeing its successful completion is indicated by the geometry of the shape enclosing that step. Steps enclosed within parallelograms, the exam steering committee. Steps enclosed by octagons, the scoring teams. Specific work products are enclosed by rounded shapes.

SUSTAINABILITY OF THE CERTIFICATION EXAM PROCESS

To ensure the sustainability of the ASBMB certification exam, we have identified several priorities:

- Expanding the community of volunteer contributors
- Growing the question bank
- Increasing the flexibility of exam administration through online delivery

Addressing these goals will allow the exam to better serve the growing number of accredited BMB programs with their

associated students and educators into the future. Continued volunteer participation, assisted by future improvements in exam-scoring software, will be essential to sustaining the exam as an accessible, high-quality assessment tool. It is noteworthy, therefore, that more than half of the current scorers have served in this role for two or more years, with approximately a third of scorers participating for at least 4 years. The community of scientist-educators affiliated with ASBMB's accreditation program thus shows tangible signs of long-term sustainability as evidenced by a stable core membership complemented by consistent leadership and continual growth (Supplemental Material 1).

Volunteer support will also be critical to expand the bank of questions for the long-term success of this dynamic exam. Maintaining an adequate question bank for each concept/skill area and level will require workshops and working groups such as those described earlier to write and refine new questions. Furthermore, cataloguing questions and tracking them through piloting, revision, and use on exams are imperative as the question bank grows.

Additionally, we are implementing administrative approaches to build capacity and increase flexibility for the growing number of accredited programs (Figure 2) and students participating (Table 3) in the certification exam each year. In 2019, we launched an online registration platform in which each accredited program is provided with a unique registration site for the certification exam. Plans to administer the exam itself electronically are being implemented for 2021. This will allow automated scoring of some questions and offer scheduling flexibility for schools.

REFLECTION ON INSTRUMENT DESIGN AND FUTURE DIRECTIONS

Social science research and discipline-based education research rely on

well-established standards to develop assessments that are relevant, fair, and beneficial to stakeholders (AERA *et al.*, 2014; Bandalos, 2018). The ASBMB certification exam arose organically from the interests of a community of BMB educators and was developed to meet immediate needs of the newly launched ASBMB accreditation program (Del Gaizo Moore *et al.*, 2018); consequently, this exam aligns well with some aspects of the accepted testing standards and diverges from others. As is often the case, our understanding of the meaning of test results and of how well the test functions to measure targeted constructs evolves over time, as more evidence is collected about the test

TABLE 3. Number of students earning certification per year

Year	Participating programs	Participating students	Certified	Certified with distinction
2014	5	193	67 (35%)	n.a. ^a
2015	27	465	194 (42%)	62 (13%)
2016	43	637	232 (36%)	65 (10%)
2017	51	664	367 (55%)	122 (18%)
2018	64	994	417 (42%)	122 (12%)
2019	73	993	412 (41.5%)	114 (11.5%)

^aThe certified-with-distinction classification was not implemented until 2015.

itself and about the relationship between testing results and relevant outcomes (Messick, 1986; Reeves and Marbach-Ad, 2016). The following section describes ways in which the exam development process aligned with standards, ways in which it differed, and plans to collect a wider range of validity evidence in the future.

A community of BMB education experts has developed the certification exam using an iterative process that recognizes BMB as a discipline and seeks to address the needs of BMB students and educators. At the outset, the community clearly defined the purpose of the exam and identified the domain of the construct to be measured. It was determined that an exam that met community needs did not already exist and that the most appropriate item format would be a mix of multiple-choice/multiple-select and constructed-response questions. A test blueprint was designed around the four core concept and skill areas previously defined by the larger BMB education community and was then used to create an initial item pool. Experts iteratively conducted item review and revision, which were enhanced through simultaneous development of scoring rubrics, thus providing validity evidence based on test content. Student responses to exam questions and pilot questions were analyzed and data were used to revise questions for subsequent exams, which provided some validity evidence based on the response process. Exam implementation was standardized across diverse institutions through dissemination of guidelines for administration. Uniformity in scoring was supported through creation of a scoring guide, defined processes for resolving scoring inconsistencies, and calculation of interrater reliability values.

Nevertheless, several aspects of the exam process diverged from accepted standards for test development. At first, large-scale field testing of exam questions occurred together with use of the certification exam by ASBMB-accredited programs. Thus, student response data used to inform the first rounds of revision were taken from exam responses that also determined whether students earned certification. Now, however, all new questions are piloted, and piloting is separate from certification. An additional piece of validity evidence not initially collected would have been think-aloud interviews as a follow-up to the response process. To date, we have also not collected validity evidence based on internal structure, relation to other variables, or consequences of testing. This is due largely to the complexity of collecting such data and the heavy reliance of the exam enterprise on faculty volunteers, who receive no compensation and only nominal professional recognition for their work. Perhaps unsurprisingly, it is not uncommon for tests developed by educators to use nonstandard procedures for assessing test validity

(Arjoon *et al.*, 2013). Although facets of validity evidence can be considered individually, crafting a convincing validity argument for a given test ultimately relies on an integrated interpretation of the evidence (Bandalos, 2018). Furthermore, as Messick asserted, test scores carry implicit value judgments. Therefore validity arguments, which define what test scores mean, are strongly tied to societal values (Messick, 1995). Given the importance of validity claims in the context of the certification exam, future directions include collecting a wider range of validity evidence in alignment with accepted standards for test development (AERA *et al.*, 2014). We identify potential types of validity evidence in the following sections, with the recognition that additional evidence will need to be considered holistically (Messick, 1995).

Validity Evidence Related to Response Process

This type of validity evidence reveals information about the construct being measured and the detailed response of the test taker (AERA *et al.*, 2014). Cognitive interviews are often considered the “gold standard,” because they can reveal whether the “psychological processes and cognitive operations performed by the respondents actually match those delineated in the test specifications” (Padilla and Benítez, 2014, p. 141). Embedding cognitive interviews with students as part of the question development process is an essential next step for investigating whether the cognitive processes used by students while answering questions align with those expected by exam developers. Moreover, moving to an online exam format may allow for monitoring of students’ response times, a related measure that correlates with the complexity of the cognitive processing of the respondent (Sireci *et al.*, 2008).

Validity Evidence Related to Internal Structure

Although the certification exam is based on four concept and skill areas, the areas are broad enough that confirmatory factor analysis may not provide interpretable validity evidence. However, the exam is structured such that we have a record of discrete characteristics of the items (e.g., difficulty and cognitive level) that would be needed to construct a Rasch model to facilitate predictions of how students will perform, manifest in the actual student performance data (Reeves and Marbach-Ad, 2016).

Validity Evidence Based on Relation to Other Variables

The certification exam is designed to assess students’ proficiency in core concepts and skills as they near completion of a biochemistry and/or molecular biology major. Therefore, it will be informative to investigate whether student performance on

TABLE 4. The 2019 exam blueprint, including questions by concept area, type, Bloom's taxonomy level, item difficulty (as indicated by the mean student score), and item discrimination

Concept area	Question number	Question type	Bloom's category	Item difficulty	Item discrimination	Quality ^a
Energy and metabolism	Q1	Constructed response	LOCS	1.71	0.355	Good
	Q2	Constructed response	HOCS	2.23	0.441	Excellent
	Q3	Multiple select	HOCS	1.74	0.223	Fair
Macromolecular structure, function, and assembly	Q4	Multiple select	LOCS	2.09	0.24	Fair
	Q5	Constructed response	HOCS	2.09	0.458	Excellent
	Q6	Multiple select	HOCS	2.24	0.229	Fair
Information storage and transfer	Q7	Constructed response	HOCS	1.64	0.304	Good
	Q8	Multiple select	LOCS	1.94	0.384	Good
Scientific method, analytical and quantitative reasoning	Q9	Multiple select	HOCS	2.51	0.324	Good
	Q10	Constructed response	LOCS	2.37	0.347	Good
	Q11	Constructed response (calculation)	HOCS	2.21	0.395	Good

^aQuality in terms of ability of the individual question to distinguish between students who scored low or high on the exam overall.

the certification exam correlates positively with successful completion of ASBMB-accredited degree programs. In the future, we plan to partner with participating institutions to identify metrics of student success in their degree programs and investigate the relationship between these metrics and performance on the certification exam. Such metrics could include cumulative grade point average in BMB courses, scores on capstone projects, and scores on key course-based assessments. Although it is possible to consider comparing performance on the ASBMB certification exam to performance on the ACS biochemistry exam, resource and time constraints mean that programs are unlikely to administer both exams.

Validity Evidence Based on Consequences of Testing

Because obtaining ASBMB certification could conceivably influence future educational and career opportunities, validity evidence based on the consequences of testing is especially relevant. Yet such evidence is perhaps the most difficult for a professional society like ASBMB to collect, because it requires extended coordination with students and institutions. The exam has intended benefits for both students (i.e., to demonstrate competitiveness against peers from across the nation independent of institutional prestige) and undergraduate programs (i.e., access to an independently constructed and scored instrument for assessing student achievement and program effectiveness; www.asbmb.org/education/accreditation). To begin compiling the information necessary to elucidate the actual impact of the exam, future directions include conducting surveys and interviews with students and accredited programs. For example, we need to understand the extent to which earning certification (or not) affects students' future career trajectory. Notably, lack of certification does not necessarily correspond to an absence of proficiency in all concepts and skills, particularly those like collaboration, which are difficult to assess but highly attractive to future employers. We must also be attentive to the possibility of unintended consequences, such as unforeseen bias against specific groups of students. How programs use aggregated exam data within their own institutions should be investigated as well. It is necessary, then, to implement a formal, objective, and quantitative process for evaluating the exam that is also open to the input of its stake-

holders. Overall, the nature of the ASBMB certification exam and its context must be considered when interpreting and basing decisions on exam scores, whether at the individual or the program level.

OPPORTUNITIES FOR DATA-DRIVEN IMPROVEMENT OF BMB UNDERGRADUATE PROGRAMS

In summary, the ASBMB certification exam is a dynamic assessment tool rooted in a robust consensus established by the BMB community regarding the core concepts and competencies that undergraduate students should master (Tansey *et al.*, 2013). There are many ways in which assessment drives teaching and learning (Momsen *et al.*, 2013; Hattie and Clarke, 2018). As part of a holistic evaluation, an instrument like the ASBMB certification exam is well poised to inform students and faculty about BMB disciplinary expectations and also to gauge the extent to which degree programs prepare students to become BMB scientists of the future. Student performance on the certification exam could provide faculty, curriculum chairs, administrators, and the entire BMB community a unique opportunity to reflect on the efficacy of their curricular and pedagogical choices, potentially shifting discussions about student success away from anecdotes toward data-driven reflections. Ideally, programs could use results from their own students' performance on the exam to identify gaps or redundancies in knowledge or skills and adjust curricula accordingly.

While several evidence-based instructional practices are available to support student learning (Bailey *et al.*, 2012; Haidet *et al.*, 2014; Evans *et al.*, 2016), there have been fewer tools for assessing students' proficiency, especially in BMB. The ASBMB certification exam is by design a multidimensional assessment; it addresses students' understanding of BMB core concepts and cross-disciplinary ideas, as well as the ability to apply these within context. In this regard, the ASBMB exam aligns with national calls to assess students in a way that raises disciplinary competency to the same level as conceptual understanding. For instance, the Next Generation Science Standards (National Research Council, 2013) emphasize the need for a multidimensional approach to curricular design and assessment within K–12 contexts, and this message has been extended to undergraduate STEM (Laverty *et al.*, 2016).

CLOSING THOUGHTS

In many ways, the ASBMB certification examination for undergraduate BMB majors represents a novel synergy between a professional society and the community that it serves. The exam and the accreditation program from which it is derived were initiated and are now powered by a team of volunteer scientist-educators informed by the input of several hundreds of their colleagues through their continued attendance at ASBMB-sponsored conferences, workshops, and webinars. While the origins and form of the exam remain largely grassroots in nature, the society provides several key ingredients. These include the imprimatur of a respected professional organization, the financial resources and professional staff needed to transform concepts into reality, and perhaps most importantly of all, a stable nexus for melding a large and diffuse set of scientist-educators into a cohesive, interactive community. To put it another way, the volunteers serve as the brains and heart of the enterprise, while the society provides the bones and sinew. Beyond the benefits of the exam itself, perhaps the most remarkable aspect of the certification exam has been the manner in which its cadre of volunteer scientist-educators has developed into a spontaneously self-improving, symbiotic community of practice.

HUMAN SUBJECTS OVERSIGHT

Approval for the accreditation program and exam (FASEB-PHSC-13-01) and for analyzing de-identified student exam responses (FASEB-PHSC-16-01) was received from the FASEB Protection of Human Subjects Committee, which determined that the study proposals meet all qualifications for Institutional Review Board exemption per the Health and Human Services regulations at 45 CFR 46.101(b).

ACKNOWLEDGMENTS

This work was funded by the Teagle Foundation and the ASBMB. The authors thank the reviewers of this *Essay* for their thoughtful comments. We also thank Cheryl Bailey for her leadership as chair of the ASBMB Education and Professional Development Committee. We deeply appreciate the dedication of everyone involved in question development and exam scoring, especially ASBMB Education Fellows: Benjamin Alper, Rafael Alvarez-Gonzalez, Michele Alves-Bezerra, Ellen Anderson, Cindy Arrigo, Christina Arther, Suzanne Barbour, Ana Maria Barral, J. Ellis Bell, Jessica Bell, Paul Black, Michael Borland, Cory Brooks, Benjamin Caldwell, Kevin Callahan, Zachary Campbell, Danielle Cass, Vidya Chandrasekaran, Joseph Chihade, Brian Chiswell, Brian Cohen, Brooks Crickard, Laura Danai, Amy Danowitz, Nicole Davis, Gergana Deevska, Rebecca Dickstein, Edward Ferroni, Kirsten Fertuck, Emily Fogle, Geoffrey Ford, Kristin Fox, René Fuanta, Scott Gabriel, Allison Goldberg, Joy Goto, Thomas Goynes, David Gross, Nicholas Grosseohme, Bonnie Hall, Orla Hart, Curtis Henderson, Jennifer Hennigan, Doba Jackson, Henry Jakubowski, Sara Johnson, Carol A. Jones, Kelly Keenan, Malik Keshwani, Youngjoo Kim, Melissa Kosinski-Collins, Cheryl Kozina, Anne Kruchten, Michael Latham, Jim Lawrence, Stefanie Leacock, Watson Lees, Eric Lewis, Robley Light, Debra Martin, Betsy Martinez-Vaz, John May, Michael Mendenhall, Pamela Mertz, Florencia Meyer, Natalie Mikita, Stephen Mills, Rachel Milner, Rebecca Moen, Sarah

Mordan-McCombs, Dana Morrone, Christopher Myer, Alexis Nagengast, Amjad Nasir, Venkatesh Nemmara, Ellie Nguyen, James Nolan, Daniel O'Keefe, Amy Parente, Mary Peek, John Penniston, Joseph Provost, Aswathy Rai, Supriyo Ray, Nancy Rice, John Richardson, Karen Rippe, Jennifer Roecklein-Canfield, Niina Ronkainen, Melissa Rowland-Goldsmith, Robin Rylaarsdam, John Santalucia, Kara Sawarynski, Marcia Schilling, Kristopher Schmidt, Jessica Schrader, David Segal, Cheryl Sensibaugh, Shameka Shelby, Joshua Slee, Alyson Smith, Dheeraj Soni, Claudia B. Späni, Amy Springer, Evelyn Swain, Uma Swamy, Blair Szymczyna, Ann Taylor, Cassidy Terrell, Candace Timpte, Pam Trotter, Sonia Underwood, Melanie Van Stry, Carrie Vance, Quinn Vega, Sarah Wacker, John Weldon, Scott Witherow, Michael Wolyniak, Ann Wright, Chuan Xiao, Yujia Xu, Philip Yeagle, Laura Zapanta, Nicholas Zeringo, Xiao-Ning Zhang, and Jing Zhang. Finally, our sincere thanks to participating institutions and students for their commitment to BMB teaching and learning.

REFERENCES

- American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education: A call to action*. Washington, DC: Google Scholar
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: Google Scholar.
- American Society of Plant Biologists. (2012). *Core concepts and learning objectives in plant biology for undergraduates*. Retrieved April 18, 2021, from <https://aspb.org/education-outreach/higher-education>, CoreSource
- Arjoon, J. A., Xu, X., & Lewis, J. E. (2013). Understanding the state of the art for measurement in chemistry education research: Examining the psychometric evidence. *Journal of Chemical Education*, 90(5), 536–545. <https://doi.org/10.1021/ed3002013>
- Arneson, J. B., & Offerdahl, E. G. (2018). Visual literacy in Bloom: Using Bloom's taxonomy to support visual learning skills. *CBE—Life Sciences Education*, 17(1), ar7. <https://doi.org/10.1187/cbe.17-08-0178>, Medline
- Bailey, C. P., Minderhout, V., & Loertscher, J. (2012). Learning transferable skills in large lecture halls: Implementing a POGIL approach in biochemistry. *Biochemistry and Molecular Biology Education*, 40(1), 1–7. <https://doi.org/10.1002/bmb.20556>
- Bain, K., Bender, L., Bergeron, P., Caballero, M. D., Carmel, J. H., Duffy, E. M., ... & Matz, R. L. (2020). Characterizing college science instruction: The Three-Dimensional Learning Observation Protocol. *PLoS ONE*, 15(6), e0234640. <https://doi.org/10.1371/journal.pone.0234640>
- Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. New York: Guilford. Google Scholar
- Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York: Longmans, Green.
- Brandriet, A., Reed, J. J., & Holme, T. (2015). A historical investigation into item formats of ACS exams and their relationships to science practices. *Journal of Chemical Education*, 92(11), 1798–1806. <https://doi.org/10.1021/acs.jchemed.5b00459>
- Bretz, S. L., & Linenberger, K. J. (2012). Development of the enzyme–substrate interactions concept inventory. *Biochemistry and Molecular Biology Education*, 40(4), 229–233. <https://doi.org/10.1002/bmb.20622>, Medline
- Brownell, S. E., Freeman, S., Wenderoth, M. P., & Crowe, A. J. (2014). BioCore guide: A tool for interpreting the core concepts of Vision and Change for biology majors. *CBE—Life Sciences Education*, 13(2), 200–211. <https://doi.org/10.1187/cbe.13-12-0233>, Medline
- Couch, B. A., Wright, C. D., Freeman, S., Knight, J. K., Semsar, K., Smith, M. K., ... & Brownell, S. E. (2019). GenBio-MAPS: A programmatic assessment to measure student understanding of Vision and Change core concepts across general biology programs. *CBE—Life Sciences Education*, 18(1), ar1. <https://doi.org/10.1187/cbe.18-07-0117>

- Crowe, A., Dirks, C., & Wenderoth, M. P. (2008). Biology in Bloom: Implementing Bloom's taxonomy to enhance student learning in biology. *CBE—Life Sciences Education*, 7(4), 368–381. <https://doi.org/10.1187/cbe.08-05-0024>, Medline
- Dean, D. M., Martin, D., Carastro, L. M., Kennelly, P. J., Provost, J. J., Tansey, J. T., & Wolfson, A. J. (2018). Assessing stakeholder perceptions of the American Society for Biochemistry and Molecular Biology accreditation program for baccalaureate degrees. *Biochemistry and Molecular Biology Education*, 46(5), 464–471. <https://doi.org/10.1002/bmb.21167>
- Del Gaizo Moore, V., Loertscher, J., Dean, D. M., Bailey, C. P., Kennelly, P. J., & Wolfson, A. J. (2018). Structuring and supporting excellence in undergraduate biochemistry and molecular biology education: The ASBMB degree accreditation program. *CBE—Life Sciences Education*, 17(4), le2. <https://doi.org/10.1187/cbe.18-09-0189>
- Emenike, M. E., Schroeder, J., Murphy, K., & Holme, T. (2013). Results from a national needs assessment survey: A view of assessment efforts within chemistry departments. *Journal of Chemical Education*, 90(5), 561–567. <https://doi.org/10.1021/ed200632c>
- Evans, H. G., Heyl, D. L., & Liggit, P. (2016). Team-based learning, faculty research, and grant writing bring significant learning experiences to an undergraduate biochemistry laboratory course. *Journal of Chemical Education*, 93(6), 1027–1033. <https://doi.org/10.1021/acs.jchemed.5b00854>
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378–382. <https://doi.org/10.1037/h0031619>
- Haidet, P., Kubitz, K., & McCormack, W. T. (2014). Analysis of the team-based learning literature: TBL comes of age. *Journal on Excellence in College Teaching*, 25(3–4), 303–333. Medline
- Hattie, J., & Clarke, S. (2018). *Visible learning: Feedback*. London: Routledge. <https://doi.org/10.4324/9780429485480>, Google Scholar
- Howitt, S., Anderson, T., Costa, M., Hamilton, S., & Wright, T. (2008). A concept inventory for molecular life sciences: How will it help your teaching practice? *Australian Biochemist*, 39(3), 14–17. Retrieved April 18, 2021, from <https://espace.library.uq.edu.au/view/UQ:184474>
- Hutchings, P. (2011). From departmental to disciplinary assessment: Deepening faculty engagement. *Change: The Magazine of Higher Learning*, 43(5), 36–43. <https://doi.org/10.1080/00091383.2011.599292>
- IBM Corporation. (2019). *IBM SPSS Statistics for MAC OS, Version 26.0*. Armonk, NY.
- Kline, T. J. (2005). Classical test theory: Assumptions, equations, limitations, and item analyses. In: *Psychological testing: a practical approach to design and evaluation* (Chapter 5). Thousand Oaks, CA: Sage. <https://dx.doi.org/10.4135/9781483385693.n5>
- Lavery, J. T., Underwood, S. M., Matz, R. L., Posey, L. A., Carmel, J. H., Caballero, M. D., ... & Cooper, M. M. (2016). Characterizing college science assessments: The Three-Dimensional Learning Assessment Protocol. *PLoS ONE*, 11(9), e0162333. <https://doi.org/10.1371/journal.pone.0162333>
- Lemons, P. P., & Lemons, J. D. (2013). Questions for assessing higher-order cognitive skills: It's not just Bloom's. *CBE—Life Sciences Education*, 12(1), 47–58. <https://doi.org/10.1187/cbe.12-03-0024>
- Mattos, C., Johnson, M., White, H., Sears, D., Bailey, C., & Bell, E. (2013). Introduction: Promoting concept driven teaching strategies in biochemistry and molecular biology. In *Biochemistry and Molecular Biology Education*, 41(5), 287–288. <https://doi.org/10.1002/bmb.20726>
- Merkel, S. (2012). The development of curricular guidelines for introductory microbiology that focus on understanding. *Journal of Microbiology & Biology Education*, 13(1), 32–38. Medline, Google Scholar
- Messick, S. (1986). *The once and future issues of validity: Assessing the meaning and consequences of measurement* (pp. 86–30). Princeton, NJ: Educational Testing Service.
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 14(4), 5–8. <https://doi.org/10.1111/j.1745-3992.1995.tb00881.x>
- Middaugh, M. F. (2010). *Planning and assessment in higher education: Demonstrating institutional effectiveness*. San Francisco, CA: Jossey-Bass. Google Scholar
- Momsen, J., Offerdahl, E., Kryjevskaja, M., Montplaisir, L., Anderson, E., & Grosz, N. (2013). Using assessments to investigate and compare the nature of learning in undergraduate science courses. *CBE—Life Sciences Education*, 12(2), 239–249. <https://doi.org/10.1187/cbe.12-08-0130>
- National Research Council. (2013). *Next Generation Science Standards: For states, by states*. Washington, DC: National Academies Press. <https://doi.org/10.17226/18290>
- Padilla, J. L., & Benitez, I. (2014). Validity evidence based on response processes. *Psicothema*, 26(1), 136–144. <https://doi.org/10.7334/psicothema2013.259>
- President's Council of Advisors on Science and Technology. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Report to the President*. Washington, DC: U.S. Government Office of Science and Technology. Google Scholar
- Reeves, T. D., & Marbach-Ad, G. (2016). Contemporary test validity in theory and practice: A primer for discipline-based education researchers. *CBE—Life Sciences Education*, 15(1), rm1. <https://doi.org/10.1187/cbe.15-08-0183>
- Schroeder, J., Murphy, K. L., & Holme, T. A. (2012). Investigating factors that influence item performance on ACS exams. *Journal of Chemical Education*, 89(3), 346–350. Retrieved April 18, 2021, from <https://pubs.acs.org/doi/10.1021/ed101175f>
- Shi, J., Wood, W. B., Martin, J. M., Guild, N. A., Vicens, Q., & Knight, J. K. (2010). A diagnostic assessment for introductory molecular and cell biology. *CBE—Life Sciences Education*, 9(4), 453–461. <https://doi.org/10.1187/cbe.10-04-0055>
- Sireci, S. G., Han, K. T., & Wells, C. S. (2008). Methods for evaluating the validity of test scores for English language learners. *Educational Assessment*, 13(2–3), 108–131. <https://doi.org/10.1080/10627190802394255>
- Smith, J. I., Combs, E. D., Nagami, P. H., Alto, V. M., Goh, H. G., Gourdet, M. A., ... & Tanner, K. D. (2013). Development of the biology card sorting task to measure conceptual expertise in biology. *CBE—Life Sciences Education*, 12(4), 628–644. <https://doi.org/10.1187/cbe.13-05-0096>
- Smith, M. K., Wood, W. B., & Knight, J. K. (2008). The Genetics Concept Assessment: A new concept inventory for gauging student understanding of genetics. *CBE—Life Sciences Education*, 7(4), 422–430. <https://doi.org/10.1187/cbe.08-08-0045>
- Tansey, J. T., Baird, T. Jr., Cox, M. M., Fox, K. M., Knight, J., Sears, D., & Bell, E. (2013). Foundational concepts and underlying theories for majors in biochemistry and molecular biology. *Biochemistry and Molecular Biology Education*, 41(5), 289–296. <https://doi.org/10.1002/bmb.20727>, Medline
- Villafañe, S. M., Bailey, C. B., Loertscher, J., Minderhout, V., & Lewis, J. E. (2011). Development and analysis of an instrument to assess student understanding of foundational concepts prior to biochemistry coursework. *Biochemistry and Molecular Biology Education*, 39(2), 102–109. <https://doi.org/10.1002/bmb.20464>, Medline
- Villafañe, S. M., Heyen, B. J., Lewis, J. E., Loertscher, J., Minderhout, V., & Murray, T. A. (2016). Design and testing of an assessment instrument to measure understanding of protein structure and enzyme inhibition in a new context. *Biochemistry and Molecular Biology Education*, 44(2), 179–190. <https://doi.org/10.1002/bmb.20931>, Medline
- White, H. B., Benore, M. A., Sumter, T. F., Caldwell, B. D., & Bell, E. (2013). What skills should students of undergraduate biochemistry and molecular biology programs have upon graduation? *Biochemistry and Molecular Biology Education*, 41(5), 297–301. <https://doi.org/10.1002/bmb.20729>, Medline
- Wiggins, G. P., & McTighe, J. (2005). *Understanding by design* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
- Wright, A., Provost, J., Roecklin-Canfield, J. A., & Bell, E. (2013). Essential concepts and underlying theories from physics, chemistry, and mathematics for biochemistry and molecular biology majors. *Biochemistry and Molecular Biology Education*, 41(5), 302–308. <https://doi.org/10.1002/bmb.20728>, Medline
- Xu, X., Lewis, J. E., Loertscher, J., Minderhout, V., & Tienson, H. L. (2017). Small changes: Using assessment to direct instructional practices in large-enrollment biochemistry courses. *CBE—Life Sciences Education*, 16(1), ar7. <https://doi.org/10.1187/cbe.16-06-0191>
- Zoller, U. (1993). Are lecture and learning compatible? Maybe for LOCS: Unlikely for HOCS. *Journal of Chemical Education*, 70(3), 195–197. <https://doi.org/10.1021/ed070p195>