Improving consistency in large laboratory courses: a design for a standardized practical exam

Xinnian Chen, Donnasue Graesser, and Megha Sah
Department of Physiology and Neurobiology, University of Connecticut, Storrs, Connecticut

Submitted 12 August 2014; accepted in final form 9 March 2015

Chen X, Graesser D, Sah M. Improving consistency in large laboratory courses: a design for a standardized practical exam. Adv Physiol Educ 39: 102–107, 2015; doi:10.1152/advan.00109.2014.—Laboratory courses serve as important gateways to science, technology, engineering, and mathematics education. One of the challenges in assessing laboratory learning is to conduct meaningful and standardized practical exams, especially for large multisection laboratory courses. Laboratory practical exams in life sciences courses are frequently administered by asking students to move from station to station to answer questions, apply knowledge gained during laboratory experiments, interpret data, and identify various tissues and organs using various microscopic and gross specimens. This approach puts a stringent time limit on all questions regardless of the level of difficulty and also invariably increases the potential risk of cheating. To avoid potential cheating in laboratory courses with multiple sections, the setup for practical exams is often changed in some way between sections. In laboratory courses with multiple instructors or teaching assistants, practical exams may be handled inconsistently among different laboratory sections, due to differences in background knowledge, perceptions of the laboratory goals, or prior teaching experience. In this article, we describe a design for a laboratory practical exam that aims to align the assessment questions with well-defined laboratory learning objectives and improve the consistency among all laboratory sections.

A question set typically contains 3–10 different questions to allow a practical exam to assess the learning objectives for that unit. The volume of these large courses creates challenges in terms of consistency of instruction and assessment among sections (8). There is a long-standing need to provide a uniform and quality learning experience and assessment for all students in laboratory courses (5, 12). Even with careful coordination, practical exams in multisection laboratory courses are likely to be handled inconsistently among different laboratory sections. To improve this situation, we have adopted a backward design (13) when preparing the course materials, including practical exams, and standardized the practical question bank in our Human Physiology and Anatomy laboratory courses. To facilitate the administration of practical exams, an online platform is used. Specifically, we define and publish laboratory learning goals and objectives, align the formative and summative assessments with defined learning objectives, and use a standardized practical exam system using a combination of an online platform for taking the test on the Blackboard Learn learning management system (LMS) and an identical practical setup at each student’s laboratory computer station. In this design, we generate sets of equivalent questions to provide consistency in the type of questions that appear on every student’s exam while randomizing the question(s) asked from each set to individualize each student’s exam. Although the practical exam is computer based, the exam also includes hands-on questions (e.g., identification of anatomic dissections, trouble shooting of experimental setups, etc.) that are answered using the materials provided to each student.

MATERIALS AND METHODS

Using a backward design to create a practical question bank. As described by Wiggins and McTighe (13), backward design is an approach in which the instructional curriculum and assessments are created based on the desired learning outcomes (Fig. 1). In this approach, instructors start by defining the learning objectives for the unit or course and then design and use formative and summative assessments that align with the learning objectives. In our Human Physiology and Anatomy courses, chief instructors articulate six to seven learning objectives for each laboratory period. These learning objectives are published in the laboratory manual. Questions in the practical exam bank are formulated to directly address one or more defined learning objectives.

To standardize practical exams, we create a question bank for each practical exam to assess the learning objectives for that unit. The questions that address the same learning objective and share the same level of Bloom’s taxonomy are grouped together in question sets (example shown in Fig. 2) (2). Each question set contributes one question to any given practical exam. As an example, if we write a practical exam that is composed of 50 questions, we first create 50 question sets. We carefully examine all questions in each set, so that similar skills are required for students to answer the questions in that set. For example, when grouping questions that calculate nerve conduction velocity, we do not place a question that requires students to synthesize information from a graph with the one that uses words to describe similar information. For example, when grouping questions that calculate nerve conduction velocity, we do not place a question that requires students to synthesize information from a graph with the one that uses words to describe similar information (Fig. 3). A question set typically contains 3–10 different questions to allow a...
degree of randomness in each practical exam, so that each student is taking an individualized exam.

The Blackboard LMS provides a mixed array of assessment tools, ranging from true/false and multiple-choice questions to short-answer questions and essays (1). To ease the grading pressure for instructors, the most frequently implemented question types on our practical exams are multiple choice, multiple answers, and fill in the blank. Although it is challenging to write multiple-choice questions that require higher-level critical thinking skills, we are able to reach higher levels of Bloom’s taxonomy by composing practical questions that require skills to synthesize information from graphs, conduct quantitative analysis, or apply knowledge to unfamiliar situations.

Creating an identical practical setup for each station. The creation of an identical practical setup at every station allows for a standardized practical exam (Fig. 4). Materials required for practical setups may include boxes of histology slides, anatomy models, preserved specimens, experimental setups, experimental results, etc. Using these practical setups, students are able to find all materials and information needed for the completion of a randomly generated practical exam as described above. For example, at each practical station, a microscope and a box of slides (labeled as A-H) is included for students to answer questions that ask them to identify histological tissues and structures (Fig. 4B). In this case, for the same question, student A might be asked to identify slide B and student B might be asked to identify slide C (Fig. 2).

For questions related to preserved specimens that require extensive labeling, it may not be practical to include a specimen at each individual station. In these cases, students are asked to come to a communal station at the front of the room, where these specimens are located. For example, when a student taking the exam encounters a question about anatomic structures on a preserved cat, an icon of a cat cues them to come to the communal station where they will find a cat in which specific structures have been labeled with numbered pins (Fig. 4E). In this question set, student A might be asked to identify the muscles at pins 1, 4, 6, and 8 and student B might be asked to identify the muscles at pins 2, 3, 4, and 12. A typical practical exam in our course requires students to visit a communal station one to three times during the exam period.

In our courses, the laboratory classrooms have one computer station for two or three students during the regular laboratory period. Therefore, students from the same laboratory section are asked to attend the practical in subgroups, so that each station will only have one student taking the practical at a given time. Depending on the number of students and practical stations in each laboratory section, the time it takes to administer the exam for an entire laboratory section may vary. For example, a typical practical exam of our laboratory courses usually lasts 45–50 min. For those sections that have 14 students, 2 h are adequate to administer the exam for the entire laboratory section. For those sections that have 21 students, 3 h are needed for the entire laboratory section.

Administering the practical exam. Because this format may be unfamiliar to some students, a description of the exam procedures and instructions is provided to each student during the week before the exam. Students are required to read the procedures and instructions, ask any questions they might have before the exam, sign an acknowledgment that they have read the instructions, and hand the signed copy to the instructor/TA as they enter the laboratory classroom. In high-volume courses with limited laboratory space and time, this protocol saves the time it would take for instructors/TAs to deliver lengthy instructions in class and allows for more time for students to take the exam.

During the practical week, each student is assigned a time slot during his or her regular laboratory period to take the practical. Upon entering the laboratory, each student can choose any one of the identical practical stations and begin the exam. Each station includes the computer used to deliver the practical exam as well as the specimens and materials needed for the practical exam. Students use their university login information to enter the Blackboard course site. Instructors/TAs provide a password for students to log into the practical exam, and the automated timer starts when the password has been correctly entered. Once a student has completed the exam, he or she can manually submit the exam or it is submitted automatically when the allowed time has elapsed. Because practical exams are administered over several days in our course, we change the password daily.

We created a computer program that maximizes the exam window and disables all hot keys for the current existing browser or for switching between different browsers. This minimizes the chance of students accidently leaving the practical page and opening another browser during the exam.

Figure 5 shows day-by-day student performance on two practical exams, one with a focus on anatomy and the other with a focus on physiology. Since we focused on the overall trend, we did not include SEs or statistical analysis in Fig. 5. Based on the overall trend, students from different sections performed fairly similar in these exams.

Reviewing and modifying the practical exam. Our practical system should best be viewed as a three-phase process, beginning with 1) practical planning, 2) practical implementation, and completing the cycle with 3) practical review and modification. The practical exam review process is a key component in our practical system, which leads back to planning for the next iteration of the practical exam. The last phase is particularly important because it assures the practical exam to be a truly standardized approach. We view this system as a dynamic instead of static process. Each iteration of the practical generation system is related to analysis of previous practical results and is also shaped by revisions of learning objectives and new instructional approaches. The process of evolution involves learning, modifying, and ultimately strengthening the overall learning process as well as the practical exam itself.

The Blackboard LMS allows instructors to run an item analysis after the administration of the practical exam (1a). The item analysis provides statistics on overall test performance as well as individual test questions. For each question, the item analysis determines the difficulty level for that question as well as a discrimination factor, an indicator for item effectiveness. The item analysis assigns the difficulty level as easy, medium, or hard, based on the percentage of students who answered the question correctly (>80%, 30–80%, and <30%, respectively). We typically include questions at all three difficulty levels on the practical exams to address the various course learning objectives. However, it is important to confirm a similar difficulty level for all questions within a given question set. If outliers
are identified, they should be reviewed for revision or omission from the exam.

The item analysis also assigns a discrimination factor (−1.0 to +1.0) to each question, which can be used to indicate the effectiveness of the question to differentiate between students with higher and lower levels of knowledge of the exam content. Questions with a discrimination score of <0.1 or negative values are flagged for review. A poor discrimination score may be due to miskeying of the answer choices, ambiguous wording of the question, or weak alignment with learning objectives. These problems can be remedied by eliminating the question, revising the question or answer choices, updating the learning objectives, or refining the course instruction.

For example, on an early version of the exam, one particular histology identification question was flagged for review and rated “hard” by the item analysis, whereas the other questions in this set were rated medium-easy. In this particular case, we found that some of the slide boxes provided to the students at the individual stations contained a slide that was not representative of the typical view of this tissue. Replacing this slide in those boxes improved student performance and made the difficulty score for this question comparable with the other questions in the same set.

One example in which we used postexam item analysis to refine instruction is a question set that assessed the following learning objective: “Describe and explain the effects of muscle stretch on the strength of muscle contraction.” In one question set aligned to this learning objective, students were asked to identify specific points on the muscle length-tension curve. One question in the set asked which point on the curve corresponds to the longest physiological muscle

Fig. 2. Example of a representative question set from a question bank. Only one question from the group is chosen for an individual test, allowing for different exams among the test takers. All questions within the question set share the same Bloom’s taxonomy level.

Fig. 3. Example of two questions that do not belong to the same question set. While the content of these two questions is similar, the skill set required to answer each one is different. In question 1, students use a formula to answer a word problem. In question 2, students must synthesize information from a physiological tracing and apply that information to the formula.

Question 1: When calculating the earthworm median giant fiber conduction velocity using the “difference method” experiment, delta d is 20 mm. The latency period for distance 1 is 5 ms and the latency period for distance 2 is 6.5 ms. The conduction velocity is:
A. 4.0 m/s
B. 13.3 m/s
C. 30.5 m/s
D. 100.0 m/s
E. 200.0 m/s

Question 2: For the “difference method” experiment, delta d is 20 mm. Use the displayed experimental recording to calculate the conduction velocity of the earthworm median giant fiber.
A. 2.0 m/s
B. 6.5 m/s
C. 10.0 m/s
D. 20.5 m/s
E. 40.0 m/s
length. For this question, item analysis revealed a difficulty level of hard and a discrimination factor of poor. A review of instruction of this concept revealed that the instructors were teaching the effects of stretch at the sarcomere level, including optimal overlap of myofilaments, and not translating these effects to the in vivo physiological constraints. In the following year, we engaged students in a more thorough discussion of the in vivo whole muscle physiology of the length-tension curve. In the next iteration of the exam, item analysis of the same question revealed a medium difficulty level with a discrimination factor of fair, indicating an improvement in student performance and equivalency in the question set.

Assessing students at appropriate Bloom’s taxonomy levels. We strive to assess a balance of factual knowledge, technical skills, and higher-order critical thinking skills. We have specifically created questions at varying levels of Bloom’s taxonomy to assess student mastery of different skills. For example, in one question set, students identify specific muscles on a model, whereas in another question set, students evaluate what limitations in movement might occur if a certain muscle (or group of muscles) is damaged. After each practical, the instructors and laboratory coordinator identify questions that would be beneficial to elevate to higher Bloom’s taxonomy levels for the next iteration of the exam.

This study was approved by the Institutional Review Board of University of Connecticut.

DISCUSSION

The use of computer-aided assessment in both formative self-tests and summative exams has been carefully discussed (6, 7, 11). Although it does not provide a universal solution for all assessments, the development of online teaching tools allows the creation of question sets and provides a convenient way to standardize an exam. The introduction of a generalized setup that allows the students to manipulate specimens, in addition to a computer-based testing tool, allows the creation of a standardized and hands-on practical exam for large courses with multiple sections.

In comparison with traditional station-to-station practical exams, this system offers several benefits to students while taking and reviewing the exam. Students who take the standardized practical exam are able to self-pace themselves rather than being strictly timed at each individual station. If they are confident in the
How We Teach: Classroom And Laboratory Research Projects

106 THE DESIGN OF A STANDARDIZED PRACTICAL EXAM

answer to a question, they can immediately move on, whereas if they are unsure of their answer they can return and review the question. In traditional practical exams, students must spend the allotted time at each station and often waste time waiting to move to the next station or, alternatively, are rushed through a station where they need more time to think. Ultimately, the ability for students to pace themselves during the exam gives instructors the freedom to ask more complicated questions that require more time for higher-order thinking.

We have found that the “view attempt” review feature in Blackboard allows students to more easily review their practical exams with instructors/TAs. The review feature provides a clear visual representation of which questions a student answered correctly and incorrectly along with all possible answer choices with the correct answer choice indicated. Traditionally, it is challenging to review station-to-station practical exams with students once the practical is broken down at the end of the practical exam period. The Blackboard review feature provides valuable feedback to both students and instructors at any time after the exam is completed.

Instructors can review overall class statistics, section statistics, and individual completed exams, at any time. This sometimes helps us to identify problematic exam questions. While we cannot change a question once students have started to take the exam, we are sometimes able to correct mistakes. For example, if we notice that a correct answer is miskeyed, we can rekey the answer choice and all student scores will be adjusted simultaneously. If we notice that a question has two possible answer choices, we can set the system to accept either answer choice retroactively. The Blackboard LMS also allows the course coordinator to monitor overall performance among different laboratory sections. We may notice that certain learning objectives are not being met by students in particular laboratory sections, and, therefore, we can identify laboratory sections that might need extra support.

This practical exam system also presents some limitations. Possible technical issues include the following. First, the exam requires the use of a computer with an uninterrupted internet connection during the course of the exam. Occasionally, technical failures at a particular station occur, and the student at that station must switch to another computer. In anticipation of these rare instances, we have a spare computer and practical setup available in each room. Second, occasionally, some students report missing images on their exam. This is due to a universal Blackboard technical issue with loading images in tests that are image intensive. The problem occurs with very low incidence and resolves itself. As a backup plan, we print a high-quality color copy of all the pictures included in an exam. If a student reports that an image is missing during the exam, the instructor can quickly provide the appropriate copy of the image to the student.

In addition to technical issues, there are time constraints to this practical design. Since two to three students usually share each computer station during the regular laboratory period, we have to create two to three different time slots to allow all students to take the practical. We also must allow for time between groups of students for students to settle down at their stations and log into the exam as well as to gather their belongings to exit the room after the exam. In addition, instructors need some time to provide direction and passwords to students as well as to recheck specimen labels between sections to ensure pins or labels have not been moved. The total time that is allowed for each practical exam has become a limiting factor, but we have found the time that we allow for each practical to be adequate to assess the defined learning objectives for each laboratory.

Time may also be a factor in the initial set up of specimens and stations. The time required to set up the practical exam depends on the number of stations being prepared and other factors. Some experimental set ups are straightforward and require little time to prepare, and others may be more intricate. Sometimes, dissections are relatively simple, and there are few structures to be labeled (e.g., a bovine eye). Other times, the dissections required are more extensive and require time and delicate technique (e.g., cat blood vessels). As mentioned previously if a specimen requires very extensive labeling, it might not be practical to include that specimen at every station. To avoid excessive preparation time, we prepare only one of these specimens for each practical room, and students view that particular specimen at a communal station. Because the setup is standardized, it is much more of a team effort than our station-to-station practical exams were, and our instructors and TAs have found this design to be simpler and less time consuming for the most part.

Another minor constraint to this practical design is availability of materials. We prepare 28 stations for each exam, and, in occasional cases, we do not have enough materials to prepare 28 identical stations. For example, we may not have anatomic models (e.g., skin or a nephron) for every station, so the models are labeled and placed at the communal station in each room.

Over the course of administering practical exams in this system, we have been able to overcome the occasional technical issues and other limitations and improve the quality of the practical exams. We have created question sets characterized by defined learning objectives, Bloom’s taxonomy levels, and an equivalent difficulty level for all questions in the set. Every student encounters identical specimens and experimental setups at their practical station, yet each exam is different because the system chooses only one random question from each question set for each exam. With this design, more standardized assessments that align better with learning objectives can be achieved even for large courses with multiple sections.

ACKNOWLEDGMENTS

The authors thank the instructors, staff, and teaching assistants of our Human Physiology and Anatomy courses for the dedication and support of this work.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: X.C. conception and design of research; X.C., D.G., and M.S. performed experiments; X.C., D.G., and M.S. interpreted results of experiments; X.C., D.G., and M.S. prepared figures; X.C. and D.G. drafted manuscript; X.C., D.G., and M.S. edited and revised manuscript; X.C., D.G., and M.S. approved final version of manuscript.

REFERENCES