ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022 ASEE
Paper ID #37443

Feature requirements for online exam administration

Edward Gehringer

Ed Gehringer is a professor in the Computer Science department at North Carolina State University, and an affiliated faculty member of the Electrical and Computer Engineering Department. He received his Ph.D. from Purdue University. His research areas are advanced learning technologies and software engineering. His teaching areas are software engineering and computer architecture.

Feature requirements for online exam administration

Abstract

Over the past two decades, learning management systems (LMSs) have each developed their own software for administering quizzes and examinations online. In addition, several standalone tools like Webassign and Gradescope have similar functionality. The rapid pivot to online learning has familiarized most engineering faculty with the features and inadequacies of current tools. These fall into four categories: question types, grading options, presentation/layout issues and integrity related. This paper considers each of the categories, from the standpoint of what is currently offered and what is desired by instructors.

Introduction

Though typically associated with online courses, online exams may be used in any kind of course. They have their own benefits, such as randomized parameters and randomized ordering of questions and answers, and their own drawbacks, such as facilitation of certain kinds of cheating [JIR]. Most online exams are administered through learning-management systems (LMSs), but standalone online homework-delivery systems are also used, such as Webassign and WeBWorK. Recently, Turitin's Gradescope, which began as a platform for grading paper exams, has developed software for online administration of exams.

In this paper, we consider the features and limitations of online exams in the five LMSs with the largest market share at the beginning of the 2021–2022 academic year (Figure 1): Canvas, Blackboard, Moodle, D2L Brightspace, and Sakai, as well as certain specialized testing engines that provide functionalities not included in the more common platforms.

To determine which features were of interest, the author asked for input from three listservs: the discussion list of the POD Network, North America's largest educational development community, the ETD-L list of the ASEE Engineering Technology division, and the SIGCSE-members list for computer-science educators associated with the ACM Special Interest Group on Computer Science Education. He used the feedback to design a survey that, with the permission of the local IRB, was posted to the same lists in December 2021 and January 2022. Fifty-five responses were received, from users of all the major LMSs and many more as well. Since multiple respondents frequently disagreed about the availability of features in their platform, online documentation was consulted in case of doubt.

Question Types

The survey asked about thirteen question types:

1. True/false
2. Multiple choice
3. Checkbox (sometimes called "multiple response" or "multiple select")
4. Matching

5. Computational (numeric answer)
6. Formulaic (answer is a mathematical formula)
7. Parsons problems (rearrange items into the correct order)
8. Coding (can check to see if program code is correct)
9. Hot spot (http://tinyurl.com/hotspotq)
10. Fill in the blank
11. Short answer (sometimes called "constructed response")
12. Essay
13. File upload

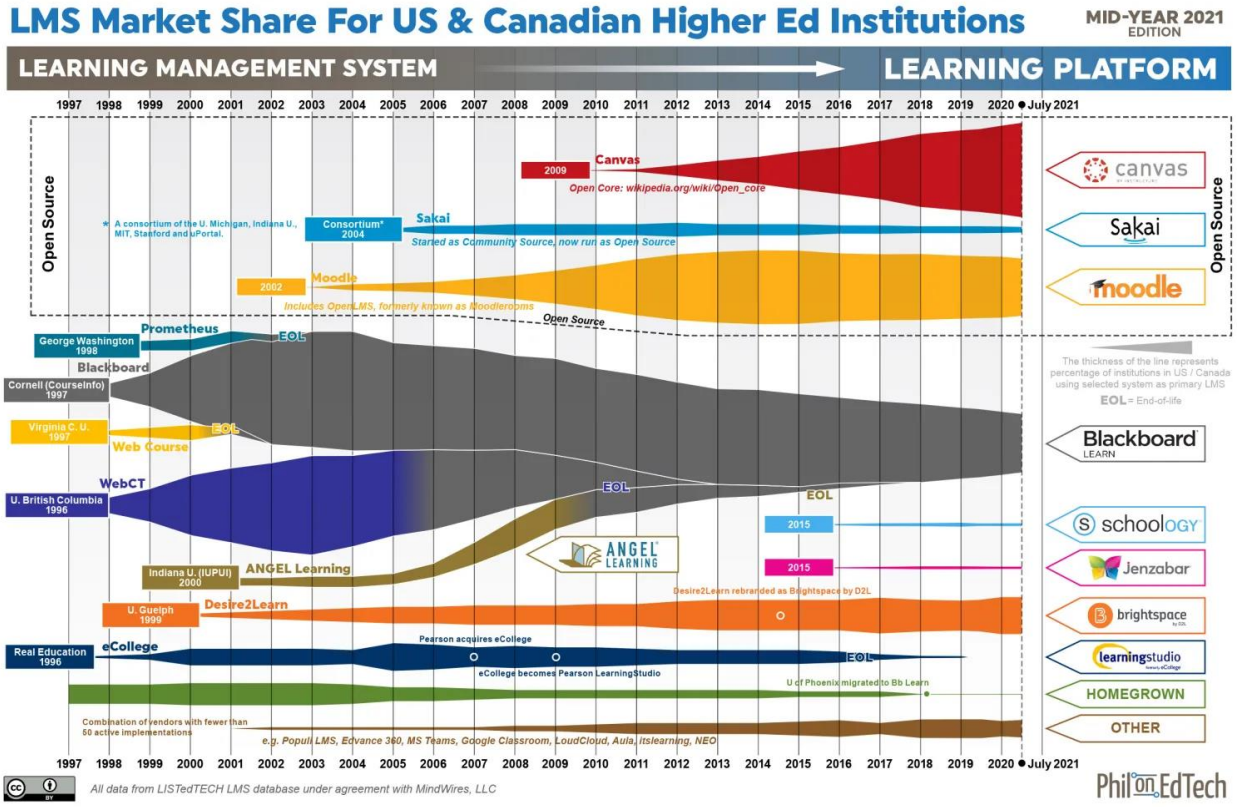

Figure 1. LMS market share of platforms over time [1]

Figure 2 shows what fraction of responding instructors rated each question type as important. The fact that the survey was posted on a listserv of computer-science instructors may have inflated the fraction of respondents who wanted coding and Parsons problems.

All systems can deliver types 1 through 5. All systems can award partial credit for checkbox questions, depending on the number of boxes that were correctly checked (or unchecked). For computational questions, it is useful to be able to specify a tolerance for numeric answers, so that any answer within a range will receive credit. In Blackboard and Sakai, this is possible,but evidently the instructor needs to specify the endpoints of the range rather than a tolerance. In Canvas, one can specify the range, a "margin" (how close the answer needs to be to the correct value in order to receive credit), or the precision required, in terms of number of digits. Moodle allows a tolerance to be specified. So does

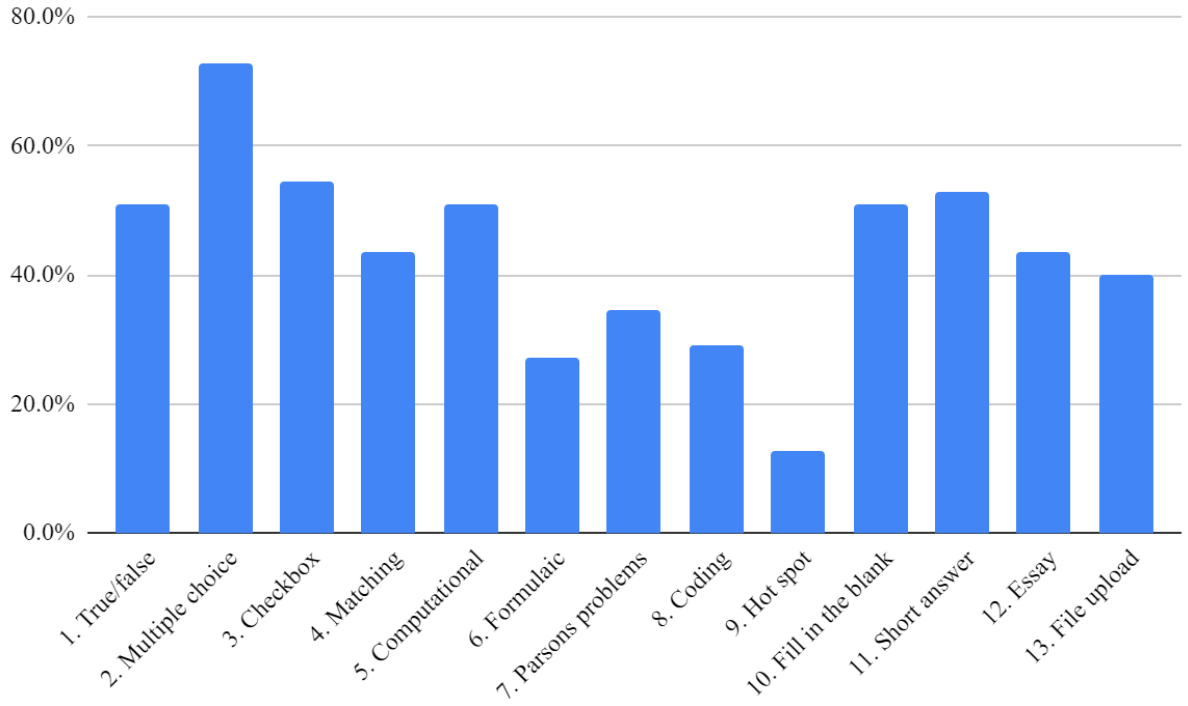

Figure 2. Fraction of instructors rating each type of question as "important"

Brightspace, but it allows the tolerance to be expressed in percentage, as well as absolute terms. In all of the five systems, the system can generate different random values for each student, so that each student gets a different numeric problem to solve.

There seemed to be a lot of confusion about question type 6, formulaic. The intent was to ask about questions that allowed a mathematical (e.g., algebraic) formula as an answer, for example, What is the first derivative of x^2? Answers like $2x$ or $x+x$ should be accepted. Several systems deliver what they call a "formula" question, but in this case, the formula is in the question, and students are asked to determine a numerical answer. As far as the author can tell, none of the top 5 LMSs deliver formulaic questions.

It also seems that none of these systems can deliver Parsons problems [2] per se. Canvas New Quizzes have an "ordering" question type. This allows students to drag various items, e.g., lines of code, into the correct order. However, New Quizzes have several limitations [3] that greatly constrain the other kinds of questions that can be asked. Blackboard and D2L Brightspace have an "ordering" question type, but it involves choosing a numeric rank for each option. This is a less friendly user interface than drag-and-drop. Parsons problems are available in the specialized programming platform Codio, which can be integrated into LMSs. Nor do any of the major LMSs support coding questions directly, though Codio does.

Another feature of Canvas New Quizzes is hotspot questions, which ask students to locate a particular position in an image. They are also available in Blackboard and Sakai.

All five systems support fill-in-the-blank questions, and all can accept multiple answers for them. Short-answer and essay questions are available in all five systems, but only Blackboard differentiates the two,

on the basis of the size of the answer blank, which is configurable for short-answer questions to indicate the length of answer expected.

File-upload questions are allowed in some format by all five systems. In Blackboard, Canvas, and Sakai, they are one of the question types. In Moodle and Brightspace, answers to essay questions can be uploaded if the appropriate option is selected.

The most-requested feature that was not provided in current platforms was regular-expression grading. For fill-in-the-blank questions, acceptable answers may take several forms: singular or plural, present or past tense, space after comma or no space after. Without regular expressions, an exceedingly large number of answer alternatives need to be provided for some questions. Regular expressions can be used in Blackboard, Moodle, and Brightspace, but not in Canvas or Sakai.

Partial Credit and Adaptive Testing

Because students don't like all-or-nothing grading, and because it does not accurately reflect their knowledge, it is very helpful to be able to automatically assign partial credit. For example, on a 3-point question, "Where was the omicron variant of COVID-19 first discovered?" the instructor could specify that the platform is to assign three points to one answer (say, "Pretoria"), two points to another answer (say, "South Africa"), and one point to a third (e.g., "Africa"), with all other answers being scored 0 points. Partial credit can be automatically assigned in Blackboard, Brightspace, and Moodle, but in Sakai, it is evidently only possible for multiple-choice questions, not for fill-in-the-blank questions. Canvas "classic" quizzes cannot autograde with partial credit. Canvas New Quizzes can, but, as noted above, fewer question types are available in New Quizzes.

Adaptive testing means that the testing engine provides questions to a student based on the student's performance on earlier questions. If the student has correctly answered several previous questions, then it would not be helpful to continue presenting "easy" questions; it makes sense to move on to harder questions. Adaptive testing is frequently used for placement tests (e.g., in language or mathematics courses); it is useful whenever it is necessary to gauge a student's skill with precision. The only one of the five platforms that implements adaptive testing is Moodle. In Moodle, an "adaptive quiz" selects questions from a question bank that accord with the apparent ability level of the student. Blackboard and Canvas have a related feature called "adaptive release." With adaptive release, a new assignment is released to a student after (s)he makes a specified score on some other assignment. For example, a student might be allowed to take Quiz 2 after scoring 90% or greater on Quiz 1. This isn't quite as flexible as adaptive testing, because it requires extra assignments to be set up and recorded in the gradebook. It prevents giving the student a single score that represents how they did on all of the adaptive questions.

Layout and Presentation

Often in testing engineering, it is useful to have multi-part questions. For example, one part of a question might ask students to calculate an answer, and the next part might ask them to explain the steps. Or they might be asked to fill in blanks in a tableau, e.g., to show which values are held in memory cells at different steps of a solution, and then to describe the process in prose. Thus, a versatile online exam

platform should be able to juxtapose several different question parts in a single question. It is not sufficient to ask the parts as consecutive *questions*, because the connections between the parts might not be obvious; and besides, when questions are delivered in random order (see the section on Combating Cheating), the various parts would be scattered randomly throughout the test.

Several of the systems provide some such functionality, but it is usually limited in ways that defeat the purpose. For example, Blackboard has a "Fill in Multiple Blanks" question, but the answer to each part needs to be a string of text (not a numeric answer, for example). In Canvas, it is possible to create multipart questions, but only if all parts are fill-in-the-blank or dropdowns. Moodle has a "Cloze" question type that can integrate almost any set of autograded question types into a single question, but free-text answers (such as explanations) are not allowed in a Cloze question. Brightspace cannot create multipart questions, but multiple questions can be grouped into a single section. Still the questions will have different numbers (e.g. Question 1, Question 2), not Question 1 (a) …. (b) … The platform that does this best, in the author's view, is Webassign, which allows "multi-mode" questions to contain up to 40 different parts of any question type that Webassign offers.

A related issue is where the blanks in the question are located. Ideally the test author should be able to place them in any desired configuration, because some questions may ask the student to fill in a table, whereas other questions are to fill in missing words in a sentence, or missing identifiers in software code. From the online documentation, it is difficult to discern how much freedom there is, and the answers of respondents to our survey do not agree. For example, of the respondents who use Blackboard, four say that the test-writer can control how question text and blanks are juxtaposed, and four say the test-writer cannot. A small plurality of Canvas respondents say the test-writer has control, and the Brightspace respondents are evenly split. Only the Moodle respondents are unanimous that the test-writer can control the placement of questions and blanks. The author can attest to this; Moodle questions can be laid out with answer blanks in any desired configuration, although the HTML editor that controls placement is not easy to use.

Combating Cheating

Academic integrity is always an issue for exams. Following the switch to online in early 2020, reports of online cheating proliferated. At Jacksonville University between 2015 and 2019, more academic-integrity cases cases were filed in the fall semester; only 47% were filed in the spring [4]. That flipped in 2020, when 61% were filed in the spring. Of those spring cases, 60% of them were brought in the last half of the semester, after the move to online learning. Moreover, three times as many of those cases involved exams (21% in 2020 vs. 7% the previous year). An even more dramatic increase was documented by the online monitoring company ProctorU [5]. From January through March of 2020, they reported cheating on fewer than 1% of exams. Then from April to June, they proctored almost four times as many exams, and detected cheating on more than 8%.

While more recent data indicates that cheating may not have increased as much online as feared, online exam platforms can and should build in support for academic integrity. One kind of support is randomization. Questions can be delivered in different orders to different students, so that students cannot refer to questions by number when surreptitiously sharing answers. Answers to multiple-choice questions can be presented in random order, to make it harder for one student to tell another which answer

to choose. Numeric questions can be displayed to different students with different parameter values, so that they cannot share answers at all. All of the major LMSs and testing platforms provide all of these functionalities.

If questions are provided to different students in different orders, then a good way to prevent cheating is to prohibit "backtracking"—students are not allowed to return to questions they have already answered. All the major platforms provide this feature, but there is concern [6] that it may disadvantage students with dyslexia or ADHD, who are often taught to answer the shorter questions first and return to longer questions later. All major platforms also integrate with lockdown browsers and support webcam monitoring. Both of those technologies have limitations and drawbacks that the author has discussed elsewhere [7].

In theory, an LMS could integrate with certain plagiarism-detection software via LTI, but neither the author nor any of the survey respondents were aware of anywhere where that was being done, save for one homegrown platform that integrates with jplag for detecting plagiarism in source code.

Several of the platforms have log data that could in principle be used to see if two students, for example, answered the same questions at nearly the same time. However, the author knows of no LMS tools that will automate this process, and in fact, Canvas explicitly states, "Quiz logs should not be used to validate academic integrity or identify occurrences of cheating." The issue with Canvas and Blackboard logs is apparently that they are not accurate enough to be relied upon in an investigation [8]. Gradescope, however, does have a feature in beta that will report pairs of students who answered questions at the most similar times. If the time of saving each answer is recorded accurately, then students who seem to answer questions in lockstep are indeed exhibiting suspicious behavior, but additional evidence, such as the number of wrong answers they gave in common, would have to be adduced to demonstrate cheating.

Observations and Recommendations

Online exam authoring and administration is a very broad field, and this study has just scratched the surface. Respondents to our survey frequently disagreed about whether their platform could deliver a particular kind of question, or group questions and answers in a specific way. This shows that instructors who use a platform are frequently not aware of all of its facilities. As long as online platforms were limited to administering low-stakes quizzes, it did not matter much whether the platform could deliver all of the kinds of questions an instructor might want to ask on an exam. But if platforms are going to administer all of the assessments for a course, then it is important for them to support the kinds of questions that an instructor would want to ask—not just because different subject areas need to ask different kinds of questions, but also because each instructor has developed a specific style of question-writing, and if a platform does not accommodate that style, then constructing new exams becomes much more work. When instructors use the same testing platform over time, the kinds of questions they ask are conditioned by the kinds of questions that are available. A move to a new testing platform may require a major change in testing strategy.

From comparing the features of existing systems with those desired by engineering faculty users, the following recommendations emerge.

Types of questions

1. Testing platforms should add support for questions that require algebraic answers.
2. Tolerances for numeric answers should be expressible in either absolute or percentage terms.
3. Allowable answers for fill-in-the-blank questions should be specifiable by regular expressions.

Layout and presentation

1. Multi-part questions should be supported, and any question type should be able to be included in a multi-part question.
2. The question-writer should have control of how a question appears on the screen, specifically, where the answer blanks are placed.

Academic integrity

1. Question answers should be logged, and the logging should be accurate enough to be used as evidence in an academic-integrity investigation.
2. Tools should be developed to use log data to identify suspicious patterns of answering questions (e.g., one student who answers "too fast," two students answer questions "in lockstep"), realizing that such evidence can only suggest cheating, not definitively prove it.
3. Testing platforms should be integrated with cheating-detection software, both for finding plagiarized text and for detecting suspicious answer patterns on multiple-choice tests.

References

[1] Phil on Ed Tech, "State of Higher Ed LMS Market for US and Canada: Mid-Year 2021 Edition," https://philonedtech.com/state-of-higher-ed-lms-market-for-us-and-canada-mid-year-2021-edition/, accessed February 11, 2022

[2] Denny, Paul, Andrew Luxton-Reilly, and Beth Simon. "Evaluating a new exam question: Parsons problems." In Proceedings of the Fourth International Workshop on Computing Education Research, pp. 113-124. 2008.

[3] University of Minnesota Information Technology, Canvas: Understand Classic Quizzes vs. New Quizzes, https://it.umn.edu/services-technologies/how-tos/canvas-understand-classic-quizzes-vs-new [accessed February 12, 2022]

[4] Lee Ann Clements, "Plagiarism and Cheating in Response to the Pandemic," International Center for Academic Integrity, June 9, 2020, https://academicintegrity.org/blog/55- 2020/june-2020/153-plagiarism-and-cheating-in-response-to-the-pandemic [accessed June 17, 2021]

[5] Derek Newton, "Another problem with shifting education online: A rise in cheating," Washington Post, Aug. 7, 2020, https://www.washingtonpost.com/local/education/another-problem-withshifting-education-online-a-rise-in-cheating/2020/08/07/1284c9f6-d762-11ea-aff6- 220dd3a14741_story.html [accessed June 17, 2021]

[6] Sonali Raje and Shannon Stitzel, Strategies for Effective Assessments while Ensuring Academic Integrity in General Chemistry Courses during COVID-19, Journal of Chemical Education 2020 97 (9), 3436-3440, DOI: 10.1021/acs.jchemed.0c00797

[7] [self-reference … redacted]

[8] Jason Kelly, What to Do When Schools Use Canvas or Blackboard Logs to Allege Cheating, https://www.eff.org/deeplinks/2021/08/what-do-when-schools-use-canvas-or-blackboard-logs-allege-cheating [accessed February 12, 2022]