“Quizzical”: A student-authored, on-line, multiple choice question writing and learning assessment tool

C. Daniel Riggs¹, Brian Sutherland², Adon Irani², and Janice Patterson²

¹Department of Biological Sciences and ²The Centre for Teaching and Learning, University of Toronto. CANADA

Abstract
To foster engagement and active learning we developed a web based, multiple-choice question writing assignment and testing tool called Quizzical. Student authored questions were screened by teaching assistants, and approved questions were used to build a practice test bank, to which students had unlimited access. Relative to previous iterations of the course, students scored higher on exams, and end-of-term course evaluations revealed that the students highly valued this experience. We have subsequently rewritten Quizzical to include dashboards for students and instructors, enabling progress and engagement to be displayed in both graphical and tabular form. These features will allow Quizzical to be used in any context as a discipline-independent learning and testing resource.

Introduction
There are numerous factors that promote subject comprehension in university classes, but active and collaborative learning exercises are arguably two of the most effective means to foster the mastery of topics. Project based assignments and/or the generation of tools or other information that can be viewed as a resource allow students to take ownership of their work/education, and evokes positive sentiments about the learning experience. In addition, such activities promote deep learning, often providing the foundational knowledge that enables critical reasoning and analytical skills to emerge.

In large university classes, student assessment often takes place through the use of multiple-choice questions (MCQ), as exam papers can be machine scored in an expeditious and equitable fashion. An additional advantage is that MCQ exams are not demanding in terms of teaching assistants’ contracts. Some of the disadvantages include relatively long time demands for writing pedagogically sound questions, and the inability of MCQs to test higher level cogitative processes, such as creativity. Despite their widespread use and aforementioned benefits, MCQ exams also can be daunting to students, and many report that they dislike the format, as questions are sometimes viewed as confusing or ambiguous. Moreover, there is generally no provision for partial credit. There likely are two primary factors that contribute to negative sentiments about MCQs. One of these is an instructor issue, while the other relates to shortcomings of students. It is clear from analyses of exams in multiple disciplines (Dibattista and Kurzawa, 2011), as well as inspection of end of chapter MCQs in textbooks (Hansen and Dexter, 1997; Masters et al., 2001) that many instructors do not abide by the general guidelines that underpin the authoring of clear and effective questions. It is assumed that students are frustrated by flawed questions, intensifying any existing negative views they may have had on
MCQs. Fortunately, many university teaching and learning centres provide guidance to instructors, and there are excellent reviews on the proper construction of questions (e.g. Haladyna et al., 2002).

Students who do not like MCQ exams may have one of several problems. First, the succinct structure of MCQs often necessitates the use of discipline specific language or technical terms with which students may not be familiar. Second, students may opt for distractors (potential answers that are incorrect) that are true statements but which do not address the question, and/or which represent common misconceptions that previous courses have not solved for them. Put somewhat differently, students may not utilize appropriate strategies for eliminating distractors. Finally, the brevity of MCQs represents a dramatic departure from the relatively free flowing dialogue of classroom lectures and associated supporting resources, making test day a somewhat foreign experience that produces test anxiety. Research in memory and cognition has spawned a relevant encoding theory known as transfer appropriate processing, which posits that information is more readily retrieved if the cognitive processes at the point of retrieval match those that were employed during the storage of the information (Lockhart, 2002). Thus, active learning in MCQ design, and open access to a large, relevant test bank would promote cognitive processing that should increase exam performance while reducing exam associated stress.

To address both instructor and student issues in MCQ testing, and to provide our students with formative practice questions, we launched a project to engage students in an active learning experiment in a sophomore level molecular biology class at the University of Toronto. We developed a web-based learning tool that we call ‘Quizzical’, which serves as a portal for student authoring of questions, a statistical tool for assessing participation and success, and as an open test bank to provide students with practice questions in a test-like environment.

**Assignment overview and Quizzical workflow**

During the first week of class, a tutorial is given that describes the features of Quizzical. Students learn about the construction of pedagogically sound questions, the value of good distractors, and are provided with tips on authoring. This information is also provided on the course website (Blackboard in our case).

Quizzical selects students randomly to author two questions on a given topic (e.g. on lecture 1 material, which is documented in the course syllabus/Blackboard website). One of these questions is to be designed as a simple ‘recall’ question, whereas the other is to be of a more difficult ‘application’ nature. Students log into a secure university server and are presented with a template to complete for each question. The student composes the stem (the question), and provides five answers, one of which is correct, while the others serve as distractors. The student indicates the correct answer by means of a check box. Importantly, for each answer/distractor, the student is required to provide a rationale for why the answer is correct, and similarly, why each distractor is incorrect. Lastly, the student is required to provide a textbook reference, preferably in the form of a figure number. Upon submission, Quizzical routes the question to the appropriate teaching assistant (TA), who evaluates it and either provides guidance for revisions, or accepts the question for inclusion in the testbank. The TA has options for assigning scores using a pull-down menu. Quizzical is capable of communicating with Blackboard or other classroom management tools to assist in the collection of grades. After a question enters the active testing pool, it is available to students on their Quizzical homepage as
part of the testing package. Students may request a quiz on single or multiple lectures, and they receive five randomized questions from the available pool.

Once a question has been attempted, Quizzical responds to indicate that the answer was correct/incorrect, and by pressing the ‘review’ button, all answers and their justifications are shown (Figure 1).

![Figure 1: Quizzical quiz output screenshot. Note that each answer/distractor is accompanied by a justification statement, and that the figure selected by the author is also displayed such that the student can immediately see linkages. The stars on the bottom right represent the question rating (1=poor to 5=excellent) that has been assigned by the examinee.](image)

At the end of each quiz, the student is presented with a summary of their performance, in a tabular format that includes the question number, it’s type (recall or application), a text reference for the question, and their result (correct or incorrect). In addition, Quizzical displays a global summary for the student on a per lecture basis, indicating the number of questions attempted, the percent correct, and the number of questions that remain in the pool (unattempted). The latter reinforces the idea that additional practice is available, and indeed, students have unlimited access to the question bank to conduct self-testing. Quizzical randomizes questions presented in quizzes, and does not reuse a question until all available questions have been attempted. One minor, yet important feature is a countdown clock that operates during the on-line quiz. The time allotted can be altered by the instructor but the default is 2.2 minutes. (This is the average time
per question allotted for a 50 question exam in our course). By default, there is no penalty for having time expired before the completion of the quiz, and indeed the student can complete it, but this feature was added to promote time management and give the student practice under conditions that they would encounter in an actual exam.

**Dashboards and Statistical Analyses**

For all stakeholders (student, TA, and instructor), dashboards are the entry points and contain information on due dates, progress, and activities. The most extensive of the dashboards is that of the instructor, which can be used to parse information on the entire class, on an individual student, on individual questions, and on TA performance. Figure 2 shows a screenshot of the instructor’s dashboard in which global information on class activity and TA activity are shown, filtered by dates. From this mockup, one can see that the class attempted from 50-150 questions per day, averaging 60-80% correct. Similarly, the progress of TA grading is presented. From this mockup, it may be deduced that TA4 has not completed the evaluation of his/her questions from the previous week, and likewise, TA2 seems to be behind his/her peers. This feature allows the instructor to quickly detect grading problems and communicate directly to the appropriate TA.

![Dashboard Screen Shot](image)

Figure 2: Screenshot of one of the instructor dashboards where data on the class performance, and progress of the TAs is shown for a given two-week interval. The tabs at the top allow the instructor to examine and filter data for an individual student or on individual questions for advising and quality control purposes.

The tabs at the top of the page allow the instructor to view and filter data for individual students, which may be useful for advising situations, and also to view individual questions. This feature generates data on the total number of attempts and the percent success for a specific question, as well as information such as the student rating of the question and the point biserial score of the question. This can be useful in several regards. First, the percent success and the biserial score can be viewed as measures of the validity and rigor of the question, as the biserial score relates
success on the question to the overall success of students on all questions. This permits the identification of easy and difficult questions and alerts the instructor to possible issues with the question that the TA did not detect. Lastly, the student rating of the question is a measure of how students perceive the question, which detects ‘fairness’.

**A one-term experiment: Quizzical’s perceived value**

We employed Quizzical Classic, a simplified version of Quizzical that does not include many of the features of the newer version, in a one semester course in a sophomore level molecular biology course of approximately 350 students. Quizzical Classic communicated to students via email, and we experienced some minor issues with communications and data collection. Nevertheless, Quizzical Classic was very effective in promoting engagement and fostering learning, and was highly valued by students. For the 314 students who completed the course, 539 questions were generated and there were 66,880 questions attempted. For engagement analysis, we compared the number of attempts for each student to their overall class performance. A priori, we anticipated that the high achievers would attempt more questions than their peers, and in some measure, this was borne out by the analyses. Figure 3 shows the mean and median values of the number of attempts by each grade category. Although there were two expected extremes, the ‘A’ students and the ‘F’ students, we were pleased to discover that most students attempted a relatively large number of questions, particularly if the median values were compared.

![Figure 3: Quizzical engagement relative to course performance. The average number of questions attempted per grade category is given. The numbers within the data bars represent the median values. Note that with the exception of the ‘F’ students, engagement tended to be consistently high.](image)

We wished to evaluate whether Quizzical was effective in promoting learning, specifically by comparing exam scores of the 2013 class with that of previous years. Fortunately, many aspects of the course were constant: instructor, textbook, topics, exam structure, prerequisites, and class time. For the past several years, we have conducted three exams. The first term test (TT1) is a 50 MCQ exam on lectures 1-8; the second, TT2, a 50MCQ exam on lectures 9-16; and the final exam consisted of 100 MCQs that tested the remaining lectures 17-24, and had a comprehensive component (lectures 1-16 material was also included). Table I shows the mean scores for the three year period of 2011-2013.
The 2013 class was informed that some Quizzical questions might appear on exams, and in fact, these questions comprised 24% and 22% of questions for the first two term tests. The 2013 class term test scores were significantly higher than those of the two previous years, even if Quizzical questions were discounted. Indeed the average point biserial values with and without Quizzical questions were very similar (0.378 vs 0.363), indicating that the student authored questions were just as reliable as those authored by the instructor. To validate Quizzical's utility, no Quizzical questions were employed for the final exam, and in addition, numerous questions from the instructor’s established test bank were employed that were deemed to be difficult. Despite the challenging nature of this exam, students performed slightly better than the previous year’s cohort, although the improvement was not statistically significant.

Student learning and success is fostered by positive experiences that reinforce commitment. We were thus interested in determining how our class perceived Quizzical’s impact on their performance. At the University of Toronto, end-of-term course evaluations are conducted by electronic means. Instructors may not ask specific questions, but rather can choose from a large number of items to solicit information. We asked our students to evaluate Quizzical by responding to question TBI071: “Online tools, used to support course activities, contributed to my learning of the course material”. Figure 4 shows that 75 of 118 (64%) of the respondents indicated that Quizzical contributed ‘a great deal’ to their learning experience, and 84% suggested that Quizzical was a positive factor (those answering ‘mostly’ or ‘a great deal’).

Students may also provide comments on aspects of the course, and 38 of 112 written comments included specific, positive statements about Quizzical. These included sentiments such as the desire for all courses to employ Quizzical, that Quizzical was an enriching and engaging experience, and that Quizzical allowed student self assessment so that they could determine what they know and identify topics that they need to work on to master. The only negative comments centered on disagreements with TA decisions, and were minor. We conclude that Quizzical was effective in promoting engagement and enhancing the learning experience for the students.

Table I: Class performance from 2011-2013

<table>
<thead>
<tr>
<th>Year</th>
<th>TT1 score</th>
<th>TT2 score</th>
<th>Final exam score</th>
<th>Number of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>58</td>
<td>64</td>
<td>61</td>
<td>287</td>
</tr>
<tr>
<td>2012</td>
<td>59</td>
<td>64</td>
<td>58</td>
<td>296</td>
</tr>
<tr>
<td>2013</td>
<td>65/61a</td>
<td>67/66b</td>
<td>60c</td>
<td>309</td>
</tr>
</tbody>
</table>

aTerm Test 1 (TT1) contained 24% Quizzical Questions; scores of 65/61 represent the average of all questions (65) and the average discounting Quizzical questions (61).

bTT2 contained 22% Quizzical Questions; scores of 67/66 represent the average of all questions (67) and the average discounting Quizzical questions (66).

cThe 2013 final exam contained no Quizzical questions.
Figure 4: Student perception of Quizzical’s value. Students were asked to evaluate Quizzical Classic in response to item TBI071: “Online tools, used to support course activities, contributed to my learning of the course material.”

**Quizzical 2.0: an upgrade**

Based on our positive experience with Quizzical Classic, we secured funding to make Quizzical a more powerful and flexible tool. For instructors, an impediment to adopting new software is their ignorance of technical needs and the difficulties in managing administrative aspects. We created a web interface questionnaire for potential users that will assist in making Quizzical 2.0 a turnkey application. The set up page allows the instructor to partition TA assignments, set due dates, establish the rubric/grading scheme, and load graphics files (e.g. zip files from textbook publisher that contain chapter images). We upgraded Quizzical Classic with the aforementioned statistical packages and dashboards, which should enhance communication between students and TAs, allow students to track their progress, and provide the instructor and/or TA with real time data to advise students. In terms of output, questions can be exported directly to Blackboard, or to other learning management systems using QTI Lite. Similarly, scores can be directly reported to learning management systems to facilitate grade collection. We believe these features will allow Quizzical 2.0 to be easily adopted by new users.

**Discussion**

MCQs are a widely used and discipline-independent means of learning assessment and information collection. In an academic environment, they can be an economical and effective way of testing, but poorly written questions compromise the validity of the test and frustrate students. Thus, adherence to best practices in writing effective questions is requisite for accurately gauging knowledge and promoting a positive learning environment. In this regard, Quizzical prompts instructors to educate themselves about best practices for writing effective questions, and communicating this information to students assists them in authoring sound questions and in writing concisely, using discipline specific terminology. Instructors also may benefit from sharing the instructor’s role with students, creating a partnership that promotes effective learning. One final advantage for the instructor relates to ‘TA education’. In many cases, course TAs are often the instructor’s graduate student(s), and engagement of the TAs in vetting questions necessitates that they have command of what might be considered foundational knowledge of their field. This enhances the background knowledge and breadth of the graduate assistant, allowing them to more readily recognize linkages that foster higher level assessment of their own thesis projects.
Quizzical has numerous benefits for undergraduates. Students become partners in the teaching enterprise and can see their instructor, TAs, and peers as allies. With regard to the latter, our students were encouraged to develop partnerships with their peers to assist one another in authoring questions. As such, cooperative learning and discussion of course materials facilitates a deeper appreciation of the topics. Students develop metacognitive skills as they create and analyze their MCQs; such skills should transfer readily to analysis of peer and instructor generated MCQs. Indeed, a recent analysis of students’ perception of the value of practice MCQs revealed that students strongly agreed that practice tests improved their chance of success on formal examinations (Douglas et al., 2012). It is also clear that access to on-line learning tools and formative quizzes is correlated with enhanced student performance on summative examinations (Kibble, 2007; Dobson, 2008; Olczak, 2013).

Bloom’s taxonomy of learning considers different cognitive levels of engagement (Anderson and Krathwohl, 2001). Requiring students to categorize their question within this framework helps them realize that deeper learning requires deeper engagement. Furthermore, requiring students to create questions helps them develop their writing skills, engage with the course material deeply, learn discipline specific terminology, and express themselves creatively. This last opportunity is often lacking in large courses, and is difficult or even impossible to test when taking a MCQ test, but designing such questions is a practical and economically feasible way to cultivate initiative/creativity.

Our experience was very positive, and one in which the students were very pleased with the process and their learning experience, despite some early issues with encoding Quizzical Classic and with communicating our expectations. Students appreciated the open access to questions and the ability of Quizzical to provide them with self-assessment quizzes, which also gave them experience with transfer appropriate processing. We feel strongly that Quizzical 2.0, enabled with statistical packages for tracking student engagement and performance, and a question rating system that they might use as a merit indicator, will increase the worthiness of the program and further enhance student engagement.

**Practical Implications**

Quizzical 2.0 is in development and testing now, and should be available for the fall semester, 2014. We anticipate that it will be a turn key application that can be easily employed by instructors in any discipline, and will require minimal interactions with IT personnel. We also anticipate that a Blackboard building block will be available to allow Quizzical to operate seamlessly within this learning management system. For those who do not use Blackboard, Quizzical grades and questions can be exported as Excel files or via QTI lite. Lastly, a word of caution regarding TA selection. Instructors are advised to choose TAs that have a reasonably good background in the subject area; otherwise the speed of grading and the reliability of questions that move into the active testing pool may become problematic.

**References**


