USING ON-LINE QUIZZES TO HELP STUDENTS LEARN PROBABILITY AND STATISTICS

Mitchum Bock and John McColl
University of Glasgow, Glasgow, Scotland, United Kingdom
Mitchum.Bock@glasgow.ac.uk

Online quizzes can be an effective and flexible means of helping learners develop key skills in probability and statistics. Quizzes give instant feedback, to help reinforce correct understanding and eliminate fundamental errors at an early stage in learning. We will describe our experience of designing and using quizzes with non-specialist and specialist students, on several different platforms including, most recently, Moodle. We describe Model Choice, a tool that helps students identify from a brief scenario the standard family of probability distributions they should work with to solve a problem. We will emphasize key design aspects of a successful quiz system, such as the importance of giving informative feedback to the learner. Using a standard platform, such as Moodle, is likely to require some compromise on design principles but building a stand-alone system to implement ideal design choices is very time-consuming.

INTRODUCTION

Assessment is a crucial component of teaching and learning, in statistics as in all other subjects. Assessment is carried out for many different purposes. It helps motivate learners to learn and gives them opportunities to integrate and demonstrate what they have learned. The teacher has the opportunity to review and reflect on what has been learned well and what has not, and to identify what needs to be improved in future teaching in order to make the learning of this or other cohorts more secure. As summarized by Ramsden (2003), "… assessment is fundamentally about helping students to learn and teachers to learn about how best to teach them."

Giving appropriate feedback to learners about their performance in assessment tasks ought also to be an integral part of the teaching, learning and assessment process in every discipline and at every level of education. Learners who have demonstrated mastery of a topic need to have their confidence in their learning reinforced, just as learners whose performance is poorer need to know where they went wrong and what they can do to improve in the future. In recent years, much of the discussion of this matter has revolved around the distinction between summative assessment, aimed primarily at grading learning, usually at the end of a course, and formative assessment, aimed primarily at helping learners identify their strengths and weaknesses and target areas that need further work. In practice, much assessment serves both purposes, for example in-course tasks might count for a few percent of the overall mark for a course.

Evidence suggests that students are often unhappy with the feedback that they receive. For example, the National Student Survey (NSS) in the UK, commissioned annually by the Higher Education Funding Council for England, provides final-year students in higher education with an opportunity to express their opinions on the assessment and feedback they have received, as well as other aspects of their higher education experience. Each year, student responses under the heading of Assessment and Feedback are substantially less favorable than those to all the other items. The three items where students express most concern are: "Feedback on my work has been prompt"; "I have received detailed comments on my work"; "Feedback on my work has helped me clarify things I did not understand". As an example of bad feedback practice, Ramsden (2003) points out that, "Comments on multiple-choice tests - if they are given at all - are usually limited to a score indicating the proportion of right answers obtained. Students do not know which questions they have got wrong, why they are wrong or what the correct answers should be."

Based on a review of the literature, Paterson and McColl (2010) point out that the wrong kind of feedback to learners can actually be counter-productive. They suggest that there are six key features of the feedback that should be given to learners in statistics: balancing positive and negative comment; commenting in appropriate detail; providing the right amount of comment; being objective; providing feedback in a timely manner; being future oriented. Further details are given in Paterson (2008).

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

Since it can be time-consuming to provide high-quality feedback on learners' work that meets these six criteria, many teachers have attempted to utilize computers to reduce the burden. For a number of years, we have used online quizzes as one means of assessing statistics students and providing feedback to them. We have used quizzes with specialist and non-specialist students at various stages in their education. In our experience, online quizzes can be an effective and flexible means of helping learners develop key skills in probability and statistics but these benefits are only realized when quizzes and quiz systems are well designed and include facilities for giving fast, context-driven and individualized feedback to learners.

DESIGN PRINCIPLES

Online quizzes can take many different forms. The ones used with the groups of students described in this paper have all been multiple-choice, usually with one correct response, three incorrect "distractors" and a "do not know" option; this is the type of quiz that is most easily created with generic software systems such as virtual learning environments (VLEs). We have also used alternative types of items, such as matching exercises and on-screen formula completion (where the learner is required to enter intermediate values before the computer completes the solution to the problem, conditional on their input). Mixing the format of items between quizzes or even within the same quiz can be helpful in maintaining the interest of learners.

Whatever the format of the items, quizzes that assess an appropriate variety of statistical skills require tables, plots and equations to be embedded within questions and possibly the capability for learners to insert mathematical symbols and formulae as part of their response. We have discovered that setting the scene for some items (for example, about analysis of variance) can require a lot of words or pictures, so we have sometimes created sets of two or three questions that follow from each other but refer to the same introduction.

Rather than giving exactly the same test to all members of a class, we have created pools of items on each sub-topic of a quiz from which items are selected at random to create a potentially different test for every learner. This helps ensure that learners cannot simply copy from others but engage with the assessment for themselves. It also allows learners to take the quiz more than once, and still be challenged by new contexts. Others (see Hunt (2007)) have opted to create a single scenario for each test item (for example, a linear regression story) but randomly generate somewhat different parameter values or datasets at random for every learner.

As noted above, one key feature of effective feedback to learners is that it is timely. Online quizzes are attractive because they have the potential to give instant feedback; the student enters an answer which is marked right or wrong immediately. At the end of a short quiz, the student receives an overall score and is instantly confirmed in their confidence on this topic or brought to the realization that they have not yet mastered it.

The benefits to learners are much increased if, instead of just being given a right/wrong response or an overall mark, context-driven, informative feedback on each answer is provided. The student is then stimulated to reflect on why their answer was wrong or right. Providing feedback to a correct answer is more valuable than might first appear, since students can stumble on the correct answer without fully understanding why it is right, especially if the quiz is multiple-choice, or value confirmation of the thought process that led them to the right answer.

Our quiz systems give students an opportunity to act on the feedback they receive; students who give a wrong answer initially are fed back some explanation of why their answer was wrong and then allowed another attempt at the same question. This reinforces the lesson in the student's mind and emphasizes the future-oriented nature of the assessment-feedback process; lessons learned as a result of one assessment activity can be applied to other tasks in the future.

Flexibility in the delivery of online quizzes maximizes their value to the learner. We make each quiz available after the topic has been covered in class but it is then left accessible until the end of the course so that the students may attempt it again in the run-up to the final examination.

Online quizzes can also provide substantial benefit for teachers. Tables of outcomes on multiple-choice test items can be generated easily so that teachers may discover which topics appear to have been learned well and which poorly. Further tables can be drawn up to show how effectively learners were able to use the feedback given to improve their performance on test items, and teachers may improve the wording of given responses or automated feedback as a result.

Learners can be invited to comment on the learning surrounding the quiz, not just the quiz itself but also the quality and relevance of the teaching of the topic.

IMPLEMENTATION

Bespoke Delivery System vs. Generic Virtual Learning Environment

In order to implement the design principles described above a bespoke web-based system was developed within the Department of Statistics at The University of Glasgow over the period 2008-2011. Facilities were provided for secure log-in (ID and password) for students enrolled in specific courses alongside a separate staff site, also with secure login. Different levels of staff users were set up, so that some staff could only add and edit questions, whilst others could delete questions and change general system parameters (e.g. number of questions per user session). Staff with higher levels of permission could also interrogate a database of usage records, for example in order to track individuals' use of the system and investigate the difficulty of particular questions.

This bespoke system was used in several courses over several years (see below) and represented a considerable investment of time and effort to both generate questions with appropriate feedback and to build, maintain and develop the delivery software. Although the bespoke delivery system had the advantage that it could facilitate the pedagogical aims of the quiz system, its weakness lay in its dependency on programming expertise and the challenges of making it easily available to the wider teaching community. It could be argued that the bespoke system also suffered from being too adaptable to preferences of the local staff and therefore constantly under development.

Prompted by these limitations, in 2013 the content of the quiz systems (questions and feedback) was migrated to a VLE, namely the free open-source Moodle[1] system. The advantages of the generic VLE over the bespoke delivery system include:

1. Full integration of the quizzes into the virtual environment alongside all aspects of course delivery
2. Extensive functionality and control developed and maintained at no cost to end users
3. Access to an extensive online community of users and developers to advise when problems are encountered and to filter and advance ideas for further developments.

In our experience, the only design issue we had to compromise on when implementing the quizzes in Moodle was the capacity to include questions that were conditional on the preceding question. It does not seem possible to select questions at random and then to have the subsequent question chosen to link with the preceding question. This limitation requires that linked questions be rewritten so that each is self-contained.

Specialist and Service Courses

The design principles described above have been applied in the context of three undergraduate statistics courses at The University of Glasgow. One of these courses is an introductory level "service" course, i.e. it caters for a broad spectrum of students from various disciplines and with various levels of mathematical education and attainment. This course gives a broad, non-mathematical coverage of statistical material from summary statistics and plots through to topics such as multiple regression and analysis of variance. The other two courses were "specialist" as they were part of the first two years of the recommended program for students wishing to study statistics through to Honors degree level. These two specialist courses covered probability and regression respectively.

For the service course, quizzes have been written for four of the main divisions of the course (Interval Estimation, Hypothesis Testing, Regression and Categorical Data Analysis). Each quiz consists of ten questions, each corresponding to a distinct area of knowledge within the topic, and the program creates a "unique" quiz by choosing one question at random from pools of questions that explore each area. There are five multiple-choice answers, only one of which is correct by default (though the delivery systems can accommodate more than one correct answer). The user has two or three opportunities to identify the correct answer, and is given informative

feedback before subsequent attempts if he or she chooses a wrong answer. An informative summary page is also provided at the end to help the user appreciate where he or she goes wrong most often.

For the regression course, the quiz system has been used towards the end of the course to provide a revision exercise. The quiz follows the format of the service course quizzes, containing ten random questions chosen by stratified sampling to ensure a broad coverage of concepts.

For the probability course, about 85% of the questions require students to recognize, from a brief description, the family of probability distributions they should work with in order to solve a problem. The quizzes proceed as described above for the service course, but at one of three levels of difficulty, roughly equivalent to University Levels 1, 2 and Honors. The three levels differ in the range of probability models they cover and also in the type of questions asked. For instance, in Level 2 and Honors quizzes, some questions describe situations in which a named standard distribution has been rejected and the student is asked to identify the reason why. The overall coverage of probability distributions across the three levels of difficult is very extensive, including virtually every standard distribution a student will meet at undergraduate level.

Formative and Summative Assessment

The quizzes have been used for both formative and summative assessment. In the service course the quizzes were first piloted as a means of formative assessment in 2008-2009. From 2009-2012 they were further used for summative assessment after being thoroughly reviewed and updated. They were administered under controlled conditions at the end of the regular practical sessions in a computer laboratory. Once used for summative assessment, the same quizzes were made available for the students to use for formative assessment in their own time. In the specialist courses to date the quizzes have only been used for formative assessment.

In the service course four quizzes were used to provide 8% of the overall assessment marks, together with four short reports (12%), two projects (25%) and a final exam (55%). The introduction of the quizzes for summative assessment was a significant saving in staff costs as they replaced two short reports which, for a class of up to 300, represents approximately 100 hours of marking. Although the temptation to make further savings through the increased use of quizzes was strong, their use was limited to a relatively small proportion of the overall assessment. This was in part a reflection of the inherent risks associated with the reliability and security of the quizzes (especially in the earlier versions of the software) but mainly because of the conviction that the best type of feedback on assessment can be given person-to-person.

STUDENT EXPERIENCE

Throughout the process of introducing and using quizzes described above, the student experience was of particular interest. When the quizzes were used with the non-specialist class from 2008 to 2012, students were routinely invited to make free text comments at the end of each test. In all, 224 valid returns were received in that time. Of these, 17 were reflections on the individual's own performance in the test but 207 were judgments on the tests themselves: 122 (59%) positive; 76 (37%) negative; 9 (4%) mixed. Of 138 different positive comments, 78 were non-specific, 25 welcomed the feedback given and another 24 highlighted the usefulness of the tests in identifying what the student knew or did not know or what they still needed to revise. Of 85 different negative comments, 42 were complaints about the wording or marking of individual test items (which often led to changes being made immediately by the development team) and 16 described the wording of the test in general to be too long or confusing. No one complained about the general idea of having a quiz or the design choices made when implementing the quizzes.

The following individual comments from students are fairly typical.

- "The lab quiz is very useful as it tells you what you got right and what you got wrong so you know which areas to improve upon."
- "The feedback is good and also to be given a chance to answer the question for a second time."

- "These tests are very good as they give feedback if an answer is wrong and another opportunity to get the correct answer. They also cover quite a broad area of the lab material and so it is good revision for exams and further reports/projects."
- "This test was useful as it gave you the correct answers at the end so you could look over what you had got wrong and what you need to work on."
- "I like the test format. Receiving feedback on incorrect answers helps you gain the correct answer and understand where you went wrong. The explanation at the end is also useful. A worthwhile test method."
- "Helpful to review topics we have gone over and the feedback system for incorrect answers is helpful."
- "Good method as the multiple choice makes you think twice about your answer and the wording."

By interrogating the detailed logs kept of all student attempts at quizzes, the following further points were noted.

- Students used the feedback effectively (e.g. 1 correct answer first time, but 7 correct altogether).
- Students vary enormously in terms of the length of time it takes them to complete a quiz when they use it for formative assessment (e.g. Multiple Regression, 1:36 to 17:26).
- It is possible to create questions of varied levels of difficulty (e.g. Multiple Regression, 23% correct up to 88% correct).

It was noted in one of the specialist courses that while the students' engagement with the quizzes was initially poor, with only a handful of students choosing to undertake the quiz when it was first introduced, the number of attempts at the quiz increased dramatically as the final exam date approached. This illustrates that the quiz system for formative assessment has the advantage that it can be accessed when the students are most motivated to learn.

FURTHER DEVELOPMENTS AND DISCUSSION

Our aim in presenting this paper has been to generate interest in online quizzes that are designed to provide immediate, individualized, context-driven feedback to students. Although we have sometimes used quizzes for summative assessment, in the sense of generating a mark that counts towards an overall course grade, we have not compromised on the principle that feedback that provokes learning should be given to students along the way. We would see much less value in an online quiz system used for summative assessment where, to quote Ramsden (2003) again, "Students do not know which questions they have got wrong, why they are wrong or what the correct answers should be."

Whether online quizzes are created using bespoke or generic software, the creation of engaging quiz items, meaningful distractors and helpful, context-driven feedback is very time consuming for the teacher. We have described the creation of a quiz system in the context of a moderately large, academic statistics group where several colleagues contributed content and checked each other's writing over a period of several years. Many teachers of statistics work alone or in very small groups and the amount of effort required to create quizzes of this sort would be beyond the resources available in those contexts. We strongly recommend those interested in creating online quizzes to find ways of forming development teams.

One area of further development that is being explored by the authors and other colleagues is the dynamic generation of quiz content using the statistical programming language R and related packages *exams* (Zeileis et al., 2012) and *knitr* (Xie, 2013). This approach has the potential to automate the generation of large banks of quiz questions, including sophisticated graphical displays, which explore various statistical concepts using a variety of data sets and parameter values. Alongside this work, we are extending the quiz questions beyond the multiple-choice format to include the cloze format, also known as embedded answers[2]. This format asks the student to submit multiple parts of structured answers either from a drop down list of options or by

requiring them to enter numbers or text. This format allows for more stimulating questions and can be used to see if students grasp the overall process of a statistical method.

ENDNOTES

[1] "About Moodle". *Moodle.org Documentation*. docs.moodle.org/26/en/About_Moodle

[2] "Embedded Answers (Cloze) question type" *Moodle.org Documentation*. http://docs.moodle.org/26/en/Embedded_Answers_(Cloze)_question_type

ACKNOWLEDGMENTS

This work has developed over a number of years and with the input and support of many people and agencies. Although not an exhaustive list, we would like to acknowledge Katrina Paterson, Helen MacGillivray, Ewan Crawford and Bradley Payne and the funding support of the JISC/ Higher Education Academy Subject Centres' Distributed e-Learning II Program.

REFERENCES

Higher Education Funding Council for England (2013). 2013 National Student Survey summary data. Accessed via www.hefce.ac.uk/whatwedo/lt/publicinfo/nationalstudentsurvey/nationalstudentsurveydata/2013

Hunt, N. (2007). Individualized statistic coursework using spreadsheets. *Teaching Statistics, 29,* 38-43.

Paterson, K. (2008). *Good practice for formative assessment and feedback in statistics courses* (M.Sc. Thesis). Glasgow: University of Glasgow.

Paterson, K., & McColl, J. (2009). Giving useful feedback to students in statistics courses. In: D. Green (Ed.), *CETL-MSOR Conference Proceedings 2009, 7-8 Sept 2009, Open University, Milton Keynes*.

Ramsden, P. (2003). Assessing for understanding. In *Learning to teach in higher education* (2nd edition). London: RoutledgeFalmer.

Xie, Y. (2013). *Dynamic Documents with R and knitr*. London: Chapman & Hall/CRC.

Zeileis, A., Umlauf, N., & Leisch, F. (2012). *Flexible Generation of E-Learning Exams in R: Moodle Quizzes, OLAT Assessments, and Beyond.* Working Papers from Faculty of Economics and Statistics, University of Innsbruck. *eeecon.uibk.ac.at/wopec2/repec/inn/wpaper/2012-27.pdf*