Mobile-Aided Formative Assessment for an Undergraduate Biology Course

Yuhua Hu and Maurice Gallagher, Member, IEDRC

Abstract—This paper reports a technology-aided learning project where mobile phones were utilized to allow students access to formative tests throughout an undergraduate third-year Biology course for a semester. Short and diagnostic tests were set up on the University’s online assessment server to help students understand complex concepts at various stages of their study. Students could have repeated attempts at these tests anytime through either web browsers or apps on their mobile phones and other mobile devices or any computers with Internet access. Through analysis of the user statistics logged on the server and student volunteers’ video learning diaries, important insights were gained as to how students make use of continuous self-assessments and personal mobile technology to enhance their learning.

Index Terms—Biology, formative assessment, mobile technology, open access.

I. INTRODUCTION

In the past two decades many higher education institutions have been facing increasing student numbers and consequently more unfavorable staff/student ratios which are identified as likely to adversely affect the support of the students, particularly with respect to assessment and timely feedback [1]. To address this problem, computer-assisted online assessment has gained more and more attention because of its well-documented advantages: 1) students can access assessments online more flexibly—anytime anywhere with multiple attempts; 2) human examiners are poor at simple tasks that require a high level of concentration whereas computer-assisted assessors are excellent at detailed, simple tasks [2]; 3) computerized assessment can provide students with more timely feedback; and 4) they can reduce time commitments and the costs of distribution and marking, facilitating the collation and allocation of student grades, the automatic generation of statistics, and greater security of data [3].

Especially in recent years, the burgeoning mobile devices industry has augmented some of these advantages even further. Mobile technologies are advancing so rapidly and industry has augmented some of these advantages even further. Mobile-aided formative assessment for an undergraduate biology course

A. Context

The project reported in this paper was aimed at implementing online formative assessments that could be accessed through students’ personal mobile devices. The course involved was a third-year biology course on microbiology (with ~80 students) in the 2011-2012 academic year. Because of the minute size of the subject matter, the usual ‘seeing is believing’ type of experiential learning model applies poorly in this field. Knowledge is therefore gained deductively from information on microbial growth, from the impact of molecular and cellular processes, and from pathology. Unsurprisingly then, despite many practical labs, our students continually ask for more tutorials and problem solving sessions to assist their conceptualization processes. Due to highly constrained staff time for providing students with more face-to-face evaluation and feedback about their understanding and knowledge, we considered using online quizzes for students to self assess and receive instant feedback throughout the course.

B. Assessment Design Principles

To ensure maximum accessibility and effectiveness, various pedagogical and technological principles were considered during the design of the online quizzes.

Technologically, in light of the high ownership of mobile devices by students nowadays, it was decided that, whenever possible, the questions must be suitable for mobile device users as well as traditional computer users so as to provide
more channels for learning in both formal and informal environment. The University’s online testing system ‘QuestionMark Perception (QMP)’ provider has developed an app for smartphones and tablets which theoretically should allow students to access tests on their mobile devices through either the Wi-Fi points available in every university building or their 3G connections at any time and from any location. However, since no other departments in this University had tried this technology with students yet, this project served as a pilot and was to gather experience and insight about the strengths and the weaknesses of the technology itself.

Pedagogically, many factors were taken into account. Firstly, open voluntary self-assessing quizzes were implemented instead of closed compulsory tests. This was largely because this was an additional learning resource which the majority of the students would be likely to utilize in order to overcome the aforementioned difficulties they had encountered with this particular subject. We also believe that open access compels users to take responsibility for managing their own time and learning, which is a graduate attribute this University endeavors to cultivate [10]. However, as reported in Swan’s study, their students’ uptake of the online quizzes over a twelve week period was disappointing when they were completely voluntary, and consequently they had to be made a compulsory 20% of the unit’s assessment [11]. This was echoed in McAndrew’s case study where the students were motivated to take part early in the project when a minor proportion of the content in the formative assessment was to be re-used in the summative assessment [1]. Along the same line of thinking, we also encouraged usage by making students aware that a compulsory test at the end of the semester would include similar questions and the test scores would account for 17.5% of their final marks of the course.

Secondly, researches have shown that timely and reiterative self-assessment promotes student memory retention [12]. Hence, our strategy was to release different sets of quizzes at different stages while the students were receiving lectures on different topics, e.g., bacteria or viruses. They had unlimited access to each set of quizzes for 3 weeks. McAndrew’s study confirmed that having an unlimited number of formative attempts encouraged practice [1].

Thirdly, it has been observed that questions focusing primarily on demonstrating an understanding of the core of knowledge are particularly suited to online assessment [13]. Such questions happened to be what we needed to aid our students’ learning in this course. However, use of mobile devices is usually of an ad hoc nature within a very limited time when the user is ‘on-the-go’ or can only spare a short period of attention. In addition, due to the limited sizes of mobile device screens, objects such as large images and tables cannot be rendered properly. Therefore, for questions to be usable on mobile devices they have to be shorter and less demanding tasks that can be completed without requiring a heavy cognitive load [14], [15]. However, we decided not to compromise the pedagogical value of such self assessments because of the technical restraints, hence we also designed three quizzes with longer and more complex questions that could be accessed through any traditional desktop or laptop computers. To ensure better system reliability for this pilot trial of mobile-assisted QMP testing, we only used multiple choice questions for all the quizzes.

Accumulating evidence indicates that, with an appropriately designed interface, online assessment which involves frequent formative testing improves student motivation and performance [16-19]. Our design was to offer ‘packets’ of opportunity for self-testing, feedback and revision at intervals throughout the course. This should encourage ‘continuous learning’ (as opposed to the typical ‘end-of-semester cramming’), thereby providing greater opportunities for developing conceptual integration across material from the entire course. Indeed, a major goal of assessment is to train students to internalize diagnostic ‘feed-back’ and transform it into ‘feed-forward’ as a means to enhance their future learning.

C. Project Procedure

In total 16 QMP assessments were designed to be released at 4 stages (QMP1-4) of the microbiology course over 11 weeks. QMP1-3 were three sets of voluntary, self-assessing quizzes with instant feedback to each question. Each QMP set consisted of 5 quizzes, and each quiz contained 5-8 questions. At the start of the first stage, a technical training session for the students was held in a computer lab where all the students could test themselves with the pre-course knowledge revision quizzes in QMP1 as well as getting familiarized with using the online system on computers and their mobile devices. QMP2 and 3 were focused on specific course sections, whereas QMP4 was a compulsory, computer-based, summative assessment which covered the entire course material and was conducted in a computer lab. QMP1, 2 and 3 each allowed open and unlimited attempts until each ‘expired’ after a 3-week interval, to discourage end-of-course ‘exam cramming’. Students were also given a range of interchangeable options in terms of mode of access for QMP1-3: via computers, via mobile phone web browsers or via smartphone apps. In addition to the user statistics gathered by the QMP server, student volunteers were enlisted to produce video diaries which provided information on the surrounding environment at the time of access to quizzes and their (immediate) reflections about the quizzes or their learning in general. On the whole, the aim of the project was to gain preliminary insight into whether students wished to utilize brief ‘study pockets’ for self testing during their ‘personal’ time.

III. Quantitative User Data

Before the project started, a quick survey was carried out to investigate the students’ usage of mobile phones. The survey revealed that almost every student owned a mobile phone (98.9%), and the majority of the phones had web browsers (65.5%) with monthly Internet data allowances (64.4%). The percentages of smartphones that were able to use the QMP apps were also moderately high (Android phones 28.7% and iPhones 20.7%). Hence, 12 of the QMP1-3 quizzes were specifically designed for use with the QMP apps whereas the remaining 3 were more suitable for use on computer screens. The QMP server recorded the students’ access time and performance for all the quizzes, and
12 quizzes also included questions that asked students to indicate the kind of devices they used at the time, and where and why they used the particular devices. 84 students completed quizzes on a total of 1421 occasions. Each student’s total use of QMP1-3 quizzes ranged from minimum 1 time to maximum 127 times with a mean of 15 times (Table I). In 308 instances the students reported how they were accessing the quizzes (Table II). Only a minority of students used their mobile phones for the quizzes, however, among these students slightly more of them (9.7%) used the apps rather than the web browsers (8.7%).

Table III shows the three sets of quizzes had considerably varied usages from individual users. While there were always students who never made any use of some quizzes, the maximum of an individual student’s access to QMP1 was 34 times, QMP2 was 61 and QMP3 was 80 times. Fig. 1 tries to visualize the comparisons of student numbers for QMP1-3 at different access frequencies. As very few students used any one of QMP1-3 for more than 15 times, the X axis of Fig. 1 excludes those extreme outliers. As is clear in the graph, only a small number of students had not accessed QMP1, whereas the usage of QMP2 was exactly the opposite. As anticipated, near the end of the course QMP3 saw a peak of usage (Median=4.0). The graph indicates that when a student did make use of a set of QMP quizzes, s/he accessed them 1-4 times on average. Another interesting pattern is that, except for ‘0’ usage, almost at every usage frequency QMP2 attracted fewer students than the other two, forming a ‘U’ shape in the comparisons. This suggests that the students seemed to have forgotten about or were much less motivated to use these self-assessments in the middle of the semester, and some of our qualitative data later in this paper may explain this ‘U’ shape effect.

**TABLE I: OVERALL USAGE OF QUIZZES**

<table>
<thead>
<tr>
<th>No. of Students</th>
<th>Total Access</th>
<th>Min Usage/ Student</th>
<th>Max Usage/ Student</th>
<th>Mean Usage/ Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>84</td>
<td>1421 times</td>
<td>1 time</td>
<td>127 times</td>
<td>15 times</td>
</tr>
</tbody>
</table>

**TABLE II: ACCESS DEVICES**

<table>
<thead>
<tr>
<th>Mobile Phone Apps</th>
<th>Mobile Phone Web Browsers</th>
<th>Computer</th>
<th>Mobile Tablet</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.7%</td>
<td>8.7%</td>
<td>78.6%</td>
<td>1.9%</td>
<td>1.0%</td>
</tr>
</tbody>
</table>

**TABLE III: NUMBER OF ACCESS TO QUIZZES**

<table>
<thead>
<tr>
<th>Usage/Student</th>
<th>QMP1</th>
<th>QMP2</th>
<th>QMP3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>34</td>
<td>61</td>
<td>80</td>
</tr>
<tr>
<td>Median</td>
<td>1.5</td>
<td>0</td>
<td>4.0</td>
</tr>
</tbody>
</table>

**TABLE IV: LOCATIONS & REASONS FOR USING MOBILES**

<table>
<thead>
<tr>
<th>Locations</th>
<th>Reasons</th>
</tr>
</thead>
<tbody>
<tr>
<td>home</td>
<td>‘It’s half time, Marseille vs Arsenal’</td>
</tr>
<tr>
<td></td>
<td>‘Using this to ease my brain into gear to prepare to do coursework.’</td>
</tr>
<tr>
<td></td>
<td>‘Ends tomorrow’</td>
</tr>
<tr>
<td></td>
<td>‘Just found password’</td>
</tr>
<tr>
<td></td>
<td>‘Revision’</td>
</tr>
<tr>
<td></td>
<td>‘Leisure’</td>
</tr>
<tr>
<td></td>
<td>‘Bored’</td>
</tr>
<tr>
<td></td>
<td>‘Testing phone compatibility’</td>
</tr>
<tr>
<td></td>
<td>‘Just to spend my leisure time before i go for a jog.’</td>
</tr>
<tr>
<td>library</td>
<td>‘I have free time as my practical section today is relatively short.’</td>
</tr>
<tr>
<td></td>
<td>‘Tomorrow no class… so I have time today to look at the questions.’</td>
</tr>
<tr>
<td></td>
<td>‘Just want to have a try on questions.’</td>
</tr>
<tr>
<td>External public spaces, e.g., bus, train, park, hostel, chaplaincy, etc</td>
<td>‘Waiting for praying’</td>
</tr>
<tr>
<td></td>
<td>‘Waiting for a friend’</td>
</tr>
<tr>
<td></td>
<td>‘Killing time travelling on the bus’</td>
</tr>
<tr>
<td></td>
<td>‘Using time on bus for sth useful.’</td>
</tr>
<tr>
<td></td>
<td>‘Because I need to improve on my grades’</td>
</tr>
</tbody>
</table>

**TABLE V: CORRELATIONS OF QUIZ USAGE**

<table>
<thead>
<tr>
<th>QMP1 Usage</th>
<th>QMP2 Usage</th>
<th>QMP3 Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>QMP1 Usage</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>QMP2 Usage</td>
<td>0.68**</td>
<td>1.00</td>
</tr>
<tr>
<td>QMP3 Usage</td>
<td>0.25*</td>
<td>0.49**</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.001 level (2-tailed); *Correlation is significant at the 0.05 level (2-tailed); N=84**

When students used their mobile phones to access the quizzes, generally they were either in their usual study locations (e.g., home, library, lab) or on travel (e.g., bus, train). Occasionally, some students tried the self-tests in more public areas such as a park or a chaplaincy. However, all their reasons for using the quizzes pointed to a tendency to fill in unexpected or transient short periods of free time with some learning. Table IV lists some typical answers that illustrate students’ ad hoc use of the online resources while waiting for something else to happen.

Unsurprisingly, there exist strong positive correlations between usages of QMP1-3 (Table V), which means the more a student used one of the 3 QMP sets, the more s/he decided to use the other two sets as well.

Undoubtedly, the question that concerned us most was: did use of these self-assessments help students improve their learning of this subject eventually? To address this question, naturally we investigated the students’ academic achievements before and after they studied this course. The average of the marks each student had obtained for all his/her second year courses was used as his/her pre-course level, whereas the QMP4 mark was used as his/her after-course level. As we were more interested in finding out whether these quizzes had helped students improve their standing among the whole class from whichever level they had been initially, we ranked the students according to their Year 2 averages (‘Ranking of Y2_Average’) as well as their QMP4 scores (‘Ranking of QMP4’). The scatterplots between ‘Total use of QMP1-3’, ‘Ranking of Y2_Average’ and ‘Ranking of QMP4’ show some outliers and non-linear relationships between variables. Therefore, the Spearman correlation tests were used. As predicted, the students who had done well in the second year were more likely to use the quizzes
Altogether 32 learning diary videos were submitted by the 4 student volunteers. The video lengths ranged from 12 seconds to 5 minutes, with reasonable image quality. To analyze the video data, we adopted the principles of the grounded theory approach, as we did not have a predefined model to decipher the data. Therefore, we extracted the key facts when we examined the videos, which might contribute to a framework that would help us understand such data better in future. The diaries provided valuable insights into student work patterns, organization and time management, motivation, and experiences and expectations of the course. Here is a list of characteristic findings:

- Students often studied at irregular times, especially at night, hence the flexible access to the quizzes was helpful.
- Students tended to associate learning activities, small or significant, with certain fixed locations only, such as home, library or computer labs. Therefore, ‘ad hoc’ learning in non-academic environments did not occur to them very often.
- Not all students were willing to commit to voluntary self-assessment, if a direct exam link of the same question style was not apparent.
- Self-assessment dwindled at the busier times, e.g., the mid-semester, particularly with less organized students.
- Long questions, with tables or images, or those in a series, were hard to follow on the phone.
- QMP increased course topic interest and stimulated deeper thinking.
- QMP helped instigate early learning of course material and organized exam revision.
- Students would like all the quizzes to be available throughout the entire course period.

Video diaries could be a useful record of self-development, and can be useful for self-reflection and for motivating better time management.

### V. Conclusion

McAndrew concluded in his report that engaging students in formative assessment was more difficult than expected [1], which was true with our case too. On the whole, our QMP quizzes attracted a fair amount of usage from the students. Almost every one of the 84 students on our third-year course made use of the 15 voluntary self-assessing quizzes—some substantially more (e.g., 127 times) than others (e.g., once) over the 10-week period, and unsurprisingly the high achievers tended to utilize these quizzes more often than the low achievers. Nevertheless, the majority of students largely ignored this resource during the middle stage of the semester. Moreover, this study shows that, although technically the undergraduate students on this course were well equipped for ubiquitous mobile learning, most did not make use of these online quizzes through their mobile devices. They still preferred to work on computers at home, in libraries and computer labs. The small percentage of students who used mobile phones to access the quizzes were normally doing so whenever they happened to have brief periods of free time. To keep students motivated for more regular use of such resources, Wang et al. suggest that:

*A thorough orientation is needed to help students understand the scope and format of m-learning activities. Students need to get ‘hooked’ in the first class session, and they need continuous encouragement to stay involved. A reward system is necessary at the beginning but instructors should aim for students’ self-motivation afterwards [15, pp.693].*
Although an orientation session was held at the beginning of this project and the students were told that the QMP4 would be very similar to QMPI-3 and its score would count towards their final marks, there was a lack of strategy to provide continuous reminder or encouragement throughout the semester. This calls for special attention in our future iterations.

The use of the quizzes might have been beneficial in that there was a significant positive correlation between the QMPI-3 usage and students’ performance on the final test QMP4. Even though this year’s data could not verify that more frequent use of the QMPS was a direct cause of better grades, there was a rather intriguing finding which revealed that the significant positive correlation only existed among the students whose ranking standing in the class improved after the course. These students’ rankings moved up probably because of their better learning capabilities, attitudes, self-motivation, strategies for time management and coping with much increased workload in the third year, or more compatible learning styles with the course’s teaching style, than the ones whose academic performance had deteriorated. Thus, this finding may suggest two possibilities: when a student had a healthy combination of the aforementioned attributes, a) the stronger s/he was academically the more s/he tended to use the quizzes; or b) the more s/he used the quizzes the better grades s/he had achieved at the end. Therefore, at least for nearly half of the students, it was possible that more frequent use of the QMP’s helped them learn the subject more effectively.

The video diaries from four volunteers have provided useful student perceptions on the strengths and weaknesses of the questions designed for mobile phone access. For example, the length, the mathematical complexity or the inter-connections of questions were reported as affecting usability considerably. One particularly interesting point was that both the academically strong and average volunteers considered the online quizzes very helpful in instigating early learning of course material and organizing exam revision, and suggested they should be available throughout the whole semester. This would be technically easy to implement, but leaving all the quizzes accessible throughout the course might compromise the pedagogical aim of encouraging continuous and responsible learning. On the other hand, the relatively weak students regarded the quizzes as not useful because there was no direct link between the exam and the quizzes. This again illustrates very well the claim that assessment is the de facto curriculum [20].

In general, the more able and more self-motivated students seemed to enjoy the intended benefits of this extra learning resource. This is in accordance with Swan’s observation that their continuous formative e-assessment influenced the study habits of students who acknowledged that the quizzes had helped them work more consistently and more effectively over the semester [11]. In contrast, the low-achievers did not actively take advantage or benefit from this provision even when it was easily and freely accessible. Although use of the QMP’s on mobile phones was low, none of our students suggested that using their mobile devices for study activities was seen as an ‘invasion’ to their ‘personal space’. In fact, the more flexible access to such resources afforded by mobile devices and the potential emergent learning opportunities were still highly appreciated (see also Stockwell’s case study [14]). This has brought two questions to our attention. Firstly, was it the technical design defects (e.g., large tables that could not be read easily on a mobile phone screen) that had discouraged wider use among our students? If so, then the second question is: is it possible to address these design issues, or are certain disciplines perhaps inherently more compatible with some technologies than others? Conole et al. have pointed out that:

...disciplines, by their nature, emphasize some skills and ways of knowing over others (Hammond and Bennett, 2002, Oliver et al., 2007 and White and Liccardi, 2006) which in turn might lead to appropriating technologies for different purposes. ... The evidence from our data did not draw out major discipline difference of this kind, although some patterns in discipline differences were evident. ... The more ‘qualitative’, ‘textual’ or ‘visual’ the subject, the less appropriate e-assessment (in the form of binary-type multiple choice questions (MCQ)) was deemed to be [21, pp.520-521].

Bioscience subjects are often very visual and quantitative, however, the current mobile software cannot seamlessly process such elements yet even with MCQ question type alone. Hence, Levy and Kennedy’s caution, as mentioned at the start of this paper, about the actual value and effect of communications technologies in their language teaching context is very much relevant to other disciplines too. The extent to which computer-aided assessment technologies can be adapted for the teaching and learning of microbiology in particular, and how they can be partnered with mobile technologies needs to be continuously explored as these technologies are evolving constantly. Nonetheless, as Beckmann has reminded us, although ‘educators must continue to fine-tune their use of such technologies, the spotlight must remain firmly on educational intentions rather than the technology being used to deliver those intentions’[22].

REFERENCE

557


