StudySieve: supporting student-generated free-response questions

Andrew Luxton-Reilly
Computer Science Dept.
The University of Auckland
Auckland, New Zealand
andrew@cs.auckland.ac.nz

Robert Sheehan
Computer Science Dept.
The University of Auckland
Auckland, New Zealand
r.sheehan@auckland.ac.nz

Paul Denny
Computer Science Dept.
The University of Auckland
Auckland, New Zealand
paul@cs.auckland.ac.nz

Beryl Plimmer
Computer Science Dept.
The University of Auckland
Auckland, New Zealand
beryl@cs.auckland.ac.nz

Abstract

Asking students to reflect on course content and ask questions about that content has been shown to improve comprehension in numerous domains. More recently, tools have been developed to store multiple-choice questions created by students in an online repository where they can be shared, evaluated and discussed with their peers. Although benefits are reported from the use of such systems, multiple-choice questions are not suitable for all teaching contexts: many instructors prefer to use free-response questions to assess learning.

We report here on a tool specifically designed to expand the contexts in which student-generated questions can be used. StudySieve supports student-generated questions with free-response answers. We investigate the way that students use StudySieve in an introductory computing course. We categorise their questions according to the cognitive dimension of the Revised Bloom's Taxonomy, and compare them with the questions used by instructors in coursework. We find that most student-generated questions belong to the lower cognitive levels, consistent with the type of questions they must answer in laboratories and tests.

This quality assured paper appeared at the 1st annual conference of Computing and Information Technology Research and Education New Zealand (CITRENZ2010) incorporating the *23rd Annual Conference of the National Advisory Committee on Computing Qualifications*, Dunedin, New Zealand, July 6-9. Samuel Mann and Michael Verhaart (Eds).

Keywords
StudySieve, contributing student, constructive evaluation, student-generated content, peer review

Introduction

Contributing student pedagogies are ones in which students contribute to the learning of others, and value the contributions of others (Hamer et al., 2008). Technology is required to support many of these pedagogies, particularly when they are implemented in large classes where the administrative overheads may be overly burdensome (Bhalerao and Ward, 2001)

One such pedagogy is that of *constructive evaluation* (Luxton-Reilly and Denny, 2010), in which students create assessment-style questions about course content, and solutions to those questions. The questions are shared with their peers, typically using an online repository. The questions in the repository are evaluated in a peer review process, and students may use the questions as a learning resource.

Although numerous systems have been developed to support peer review (Luxton-Reilly, 2009), these more general tools lack the features required by constructive evaluation. A number of previous tools have been developed in order to support constructive evaluation in the domain of multiple choice questions (Yu, Liu and Chan, 2002; Horgen, 2007; Denny, Luxton-Reilly & Hamer, 2008). The authors of such systems report many benefits for students including increased motivation (Horgen, 2007); improved confidence (Yu et al., 2002); and improvement of content knowledge, as measured by examinations (Denny, Hanks and Simon, 2010).

Multiple choice questions can be an effective assessment and learning tool, however, it is difficult to write multiple choice questions that assess higher-order cognitive skills (Palmer and Devitt, 2006). As an example, introductory programming courses usually assess code construction tasks. Typically, this is a task such as writing a fragment of code that fulfills a given specification. This would be extraordinarily difficult (if not impossible) to assess in the framework of multiple choice questions.

To extend the domain in which constructive evaluation can be used, we have implemented a tool (entitled StudySieve) that supports the creation, evaluation and sharing of free-response questions and their answers.

StudySieve

Figure 1 shows the default view of the StudySieve system. Students are able to view existing questions, create a new question, look at their contribution statistics or find out more about the way the tool is used.

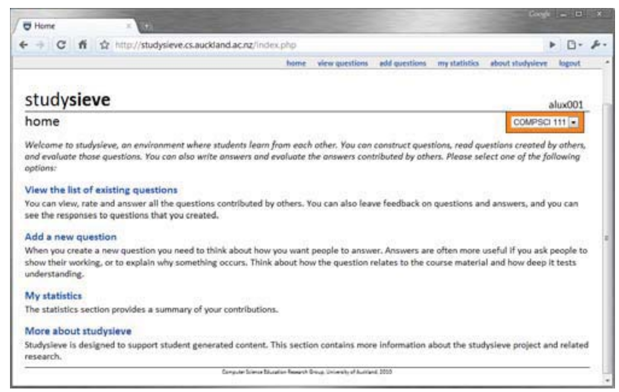

Figure 1: The StudySieve system as it appears to students when they first log in.

The list of questions may be filtered and ordered, which makes the question repository more manageable for students. Questions may be rated on a 1-5 scale, and both the user's rating and the average rating is displayed. Figure 2 shows the appearance of the question list, with the user's rating indicted by stars, and the average rating indicated by a horizontal bar directly below the stars. The third question has been rated by the user as 4 out of 5 stars, and has an average rating of about 4.5.

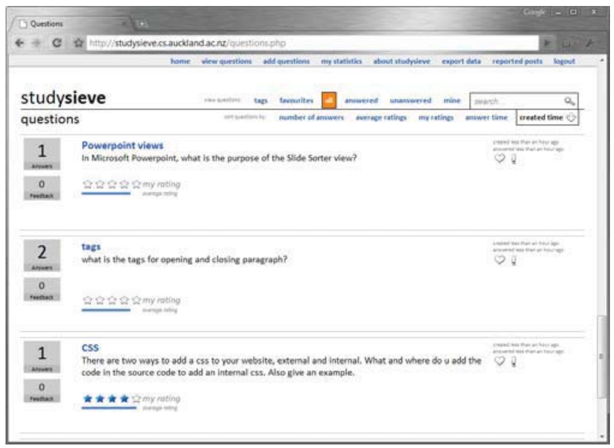

Figure 2: The questions view showing three of the student-generated questions.

The answers to a question remain hidden until a student submits an answer of their own. After the answer is submitted, the answers of other students are displayed in a list below the question. Students may rate the answers submitted by others, and can see their own ratings, and the overall average rating of the answer. Figure 3 shows a question that has been answered, and two of the answers submitted by other students.

In this study, we investigate how students in a general computing course use StudySieve, and the nature of the questions they create. We report on the logs of use, and analyse the cognitive level of the questions created by students, compared with the level of the questions used by instructors in laboratories and archived tests.

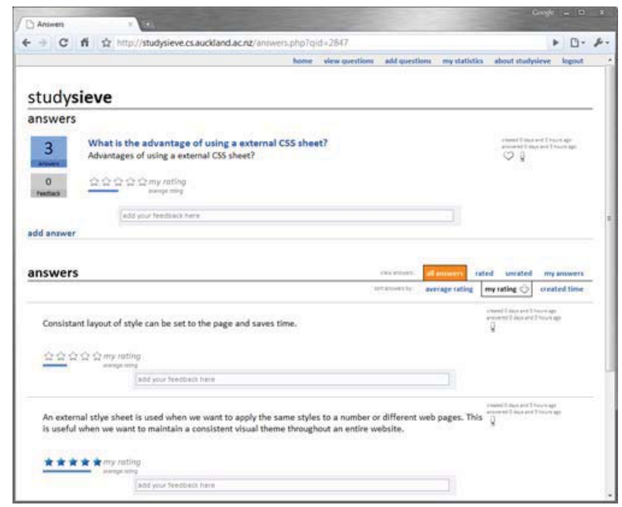

Figure 3: A question that has been answered by the user, and two answers contributed by other students.

Methodology

The StudySieve system was trialed in a first-year computing course – CS111 – entitled "An introduction to practical computing". This course is designed to provide a general overview of computing for non-CS majors. In addition to lectures, students are expected to participate in one laboratory session per week.

Students were required to submit one question to Studysieve and to answer one other question each week as part of their laboratory requirements. We report here on the data collected between 7th March 2010 and 1st April 2010, the period between the introduction of StudySieve in the first laboratory and the date of the mid-semester test.

In order to analyse the cognitive level of the questions produced by students, two of the authors coded the questions according to the Revised Bloom's Taxonomy (Anderson et al. 2001). We used only the cognitive levels dimension and coded questions as belonging to one of the categories: Remember, Understand, Apply, Analyse, Evaluate or Create. We interpret each of these categories as follows:

Remember – questions that require students to recall or remember something. This includes questions asking students to describe something, or state a definition of something.

Understand – questions that require students to explain a concept or idea using their own words. Typically this differs from simple memorization by asking students to interpret or translate ideas

Apply – questions that require the application of a process, such as performing a calculation.

Analyse – questions that require students consider how parts relate to a whole.

Evaluate – questions that require a judgment about quality based on standards and criteria.

Create – questions that require students to develop something new

Initially, a sample of 30 questions were coded by the first two authors. Once agreement was obtained, the set of questions was divided in half and questions were coded individually. Questions for which the category was ambiguous were flagged for discussion and were assigned a category after agreement was reached between the coders. After the completion of the coding a sample of 50 questions was coded independently by one of the other authors. The codes agreed in 85% of cases. The remaining questions were borderline cases falling into the categories of either Remember or Understand.

We hypothesize that the type of questions that students submit are likely to be influenced by the type of questions that instructors provide in laboratories and archived tests from previous years. We therefore categorized instructor questions present in the first three laboratories, and the three most recent archived tests, which are available to students.

Results

We report here on the usage of the StudySieve system and the cognitive level of the questions contributed by students.

Use of the system

At the time of the data collection, students had attended three laboratory sessions, one per week. The assessment requires that a laboratory report is submitted one week after a session is attended. Therefore, all students should have completed the requirements for two laboratories. Some students will have completed the requirements for the third laboratory, while others will not yet have finished the tasks specified for their third laboratory.

Questions contributed

Students were required to submit at least two questions. Some students would have submitted a third question to fulfill the requirements of their third laboratory. Students who submitted four or more questions did more than required for no additional credit. The maximum number of questions by a student was 13.

Table 1: The number of students who contribute less than, equal to and greater than the number of questions required by the assessment.

Number of questions	Number of students
0	100
1	71
2 - 3	321
>= 4	48

Answers contributed

Students were required to submit at least four answers (two for their own questions and two for questions written by others). Students who were working on their third laboratory may have submitted five or six answers. Students who submitted seven or more answers did more than required for no additional credit. The maximum number of answers by a student was 35.

Table 2: The number of students who contribute less than, equal to and greater than the number of answers required.

Number of answers	Number of students
<= 3	196
4 - 6	230
>= 7	115

Ratings contributed

StudySieve is structured so that a user must self-assess their questions and answers by assigning a rating to them when they are submitted. This means that every question and answer is rated at least once, by the student who created it. Students who have fulfilled the assessment requirements will have submitted at least three ratings for each lab (i.e. six ratings at the time of data collection). The maximum number of ratings by a single student was 196.

Table 3: The number of students who contribute less than, equal to and greater than the number of ratings required assessment.

Number of ratings	Number of students
<= 5	184
6 - 9	195
>= 10	162

Feedback contributed

Feedback may be provided on any question and on any answer. The contribution of feedback is not assessed in any way. The maximum number of comments contributed by a single student was 8.

Table 4: The number of students who contribute one or more comments as feedback.

Number of comments	Number of students
1	82
2	16
>= 3	9

SUMMARY OF USE

Overall, only 83.4% of the students enrolled in the course logged into the StudySieve system and made a contribution (i.e. submitted a question, answer, rating or feedback). Relatively few students contributed a comment as feedback to a question or answer. Table 5 shows the number of students enrolled, the number of students who logged into the system (active users), and the total amount of contribution in each category.

Table 5: Overall use of the system by students.

Category	N	Users
Students Enrolled		541
Active users		451
Questions contributed	1151	439
Answer contributed	2422	449
Ratings contributed	4078	451
Comments contributed	172	108

Cognitive level of the questions

Table 6 summarizes the number of questions authored by students in each category, and the number of instructor questions (as they appear in laboratories and archived tests) in each category. Some questions were incomprehensible or were clearly not related to course content, and as such were placed in the category of Irrelevant. The vast majority of questions contributed by students fall into the lower cognitive categories: Remember and Understand. Examples of student-generated questions in each of the categories are shown in figure 4.

Table 6: Number of questions from students and instructors that fall into each cognitive category

Cognitive level	Student questions	Instructor questions
Remember	693	54
Understand	316	21
Apply	22	14
Analyse	0	0
Evaluate	1	1
Create	0	0
Irrelevant	60	0

We identified numerous duplicate questions during the coding process. In some cases, the same student submitted the same question (or a minor variation) multiple times, in other cases multiple students submitted the same question (or a minor variation) at different times. For example, the question "What does CPU stand for?" appeared 12 times. Some questions are simple variations of the questions used in the laboratories. For example, the question "What is the difference between Cc and Bcc?" appeared in the laboratory, and 34 variations were discovered in the repository.

Discussion

We observed that a substantial portion of the class (16.4%) did not contribute anything to StudySieve. The COMPSCI 111 course is a general education course designed for students who are not majoring in CS. It is typically viewed as a low priority course by students, which may explain the lack of involvement in a straightforward assessment task.

We also note that some students used StudySieve more than the minimum required. This provides evidence that students see intrinsic value in the system. We did not observe heavy use of StudySieve in the period immediately prior to the test, as has been reported with the PeerWise tool (Denny, Luxton-Reilly & Hamer, 2008). We speculate that perhaps the large number of questions and few answers per question limited the value of the repository for revision. At the time of data collection, no facility to search the questions was available, although we have since implemented searching and tagging to assist students who are revising specific topics.

StudySieve does not require students to evaluate the questions or answers of others. Few of the questions or answers submitted by other students were rated. This is consistent with findings by Barak and Rafaeli (2004) who reported that few students were willing to assess or criticize their classmates work.

Studies conducted with peer review report that students often do not feel comfortable evaluating their peers, particularly when the criteria are imprecise (Dochy, Sergers & Sluijsmans, 1999). The rating of 1—5 stars used in StudySieve is perhaps too vague for students to confidently assess an answer. Further investigation into the evaluation of questions and answers is warranted.

The repetition of previous questions suggests that the system would be improved if there was a mechanism to either prevent or aggregate similar questions. It may be possible to extend the current interface for creating a new question so it shows similar questions that already exist in the repository before submission is confirmed. Alternatively, allowing students to identify similar questions in the repository might allow those questions (and their answers) to be combined.

REMEMBER:

- What does XHTML stand for?

UNDERSTAND:

- How can grouping shapes in Microsoft Word make it easier to manage the content in a document?

APPLY:

- Convert the binary number 010111 into a decimal number.

EVALUATE:

- Browsing "Computer", or "My Computer" in Windows XP. One will notice there may be different hard drive formats -NTFS and FAT32. What is the main difference between them and why would one user select one over the other, and what do the abbreviations stand for?

IRRELEVANT:

- QWERTY: the layout of keyboard
- It's 3:08am and you're buying a pie at the service station. What must you always do?

Figure 4: Examples of student-generated questions categorized as Remember, Understand, Apply, Evaluate and Irrelevant.

The majority of questions created by students fell into the lower levels of the Revised Bloom's Taxonomy. This is appropriate for an introductory computing course, and is consistent with the instructor questions in the laboratories and archived tests for the early topics in the course. Topics that occur later in the COMPSCI 111 course tend to involve tasks that cognitively more challenging, such as: programming in python, generating spreadsheet formulae, and constructing LaTeX for mathematical formulae. If students generate questions that have a similar cognitive level to those generated by instructors, we would expect to see a greater number of questions belonging to higher levels in the later weeks of the course.

Many of the questions that fall into the *Remember* category require a definitive answer. For example, "What does X stand for?" where X is an acronym. Questions of this type leave little opportunity to learn through the variation in answers provided by peers (since there is little or no variation in answers). For questions that require recall, a multiple choice system may be a more appropriate tool.

Conclusion and future work

StudySieve successfully supports constructive evaluation for questions with free-response answers. The questions created by students are consistent with those created by instructors for the material studied in the early part of the course. Further study is required to determine if students continue to generate similar questions to instructors when the course content is more complex and instructor questions are at a higher cognitive level.

The large number of similar questions may have occurred as a consequence of a large number of students writing questions about content that is largely factual. However, it is likely that the quality of the repository could be improved by introducing features that helped students to identify and manage repetitive questions.

Although students appear to value the repository of questions created by their peers, the majority are not engaged with the evaluation of questions. Further work is needed to improve the design of the system so students are encouraged to participate in the evaluation process.

References

Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J. and Wittrock, M.C. (Eds.) (2001) *A taxonomy for learning and teaching and assessing: A revision of Bloom's taxonomy of educational objectives.* Addison Wesley Longman, 2001.

Barak, M. & Rafaeli, S. (2004) On-line question-posing and peer-assessment as means for web-based knowledge sharing in learning. *International Journal of Human-Computer Studies 61*, 84-103.

Bhalerao, A. & Ward, A. (2001). Towards electronically assisted peer assessment: a case study. *Association for Learning Technology Journal 9*, 26-37.

Denny, P., Luxton-Reilly, A. & Hamer, J. (2008). The PeerWise system of student contributed assessment questions. In Simon & Hamilton, M. (ed.) *CRPIT 78: Proceedings of the Tenth Australasian Computing Education Conference,* 69-74: ACS

Denny, P., Hanks, B. & Simon, B. (2010) Peerwise: replication study of a student-collaborative self-testing web service in a U.S. setting. *In SIGCSE '10:*

Proceedings of the 41st ACM technical symposium on Computer science education, 421-425: ACM

Dochy, F.; Segers, M. & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education* 24, 331-350.

Hamer, J., Cutts, Q., Jackova, J., Luxton-Reilly, A., McCartney, R., Purchase, H., Riedesel, C., Saeli, M., Sanders, K. & Sheard, J. (2008). Contributing student pedagogy. *SIGCSE Bull. 40*, 194-212: ACM

Horgen, S. A. (2007). Pedagogical use of multiple choice tests — Students create their own tests. In Kefalas, P.; Sotiriadou, A.; Davies, G. & McGettrick, A. (Eds.) *Proceedings of the Informatics Education Europe II Conference*: SEERC

Luxton-Reilly, A. (2009) A systematic review of tools that support peer assessment. *Computer Science Education 19*, 209-232

Luxton-Reilly, A. & Denny, P. (2010). Constructive evaluation: a pedagogy of student-contributed assessment *Computer Science Education. In press.*

Palmer, E. J. & Devitt, P. G. (2006) Constructing multiple choice questions as a method for learning *Annals of the Academy of Medicine, 35*, 604-608

Yu, F.-Y.; Liu, Y. H. & Chan, T.-W. (2002). The Efficacy of a Web-Based Domain Independent Question-Posing and Peer Assessment Learning System *Computers in Education, International Conference on*: IEEE Computer Society