

This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Self-assessing with Adaptive Exercises

PLEASE CITE THE PUBLISHED VERSION

PUBLISHER

© Loughborough University

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Yong, Chye-Foong, and Colin Higgins. 2019. "Self-assessing with Adaptive Exercises". figshare.
<https://hdl.handle.net/2134/1974>.

SELF-ASSESSING WITH ADAPTIVE EXERCISES

Chye-Foong Yong and Colin A Higgins

Self-assessing with adaptive exercises

Chye-Foong Yong and Colin A Higgins
University of Nottingham
School of Computer Science and IT
Wollaton Road
Nottingham NG8 1BB
czf@cs.nott.ac.uk and cah@cs.nott.ac.uk

Abstract

In assessment it is usual for students to be presented with the same sets of questions, regardless of their individual ability. That is, the questions are not tailored to the students' individual performance, understanding and ability, and thus may be either too easy or too difficult for the student. However, computerised-adaptive testing (CAT) systems, present questions of appropriate difficulty for the student based on their ability, allowing the student to progress at their own pace. This paper presents such a system for self-assessment purposes and the findings from its use by a cohort of computer science students at first year undergraduate level.

Introduction

In standard computer-based test systems, the questions presented to the students are fixed for the cohort and are not tailored according to the students' individual performance during the test. As a result, students can be presented with questions that are either too easy or too difficult (Lilley and Barker, 2002; Eberts, 1997). As a result, researchers began to look into computerised-adaptive testing (CAT) systems. Such systems administer an examination with questions of appropriate difficulty for the student based on his/her ability (Syang and Dale, 1993).

A common form of adaptive testing involves a computer-administered test in which the presentation of each test item and the decision to stop the test are dynamically adapted to the student's performance in the test (Abdullah, 1999; Lilley and Barker, 2003). The most common stop conditions are:

- a certain number of questions have been administered;
- a time limit has been reached; and
- a certain standard error for the ability has been met.

The advantages of this form of adaptive testing are:

- it presents to the student only those test items which are likely to yield additional information;
- it can distinguish careless slips and guesswork from actual knowledge by repetitive testing; and
- it can arrive at an initial student model quite accurately in a relatively short time.

Notwithstanding these advantages, this form of adaptive testing has some drawbacks particularly as most CAT systems are used for summative assessment. Research carried out by (Lilley and Barker, 2002, and 2003), comments that the students would be more receptive to the use of such a CAT system in a formative rather than in a summative assessment environment. This suggests that lecturers foresee problems concerning the score method used within CAT systems. In CAT systems, the final score given to a student is typically based on the number of questions answered correctly and incorrectly, as well as on the level of difficulty of these questions. As a result, students who have answered the same number of questions correctly would almost certainly have different final scores which can bring uncertainties about the “fairness” of the assessment.

Also, in their reports, Lilley and Barker highlighted that their expert evaluators (namely some lecturers in Computer Science and in English for Academic Purposes, at the University of Hertfordshire) believed that their prototype would give greater assistance in formative rather than in a summative environment. The expert evaluators suggested that formative assessments provide the lecturers with more information regarding the students’ strengths and weaknesses, since they are typically undertaken on a regular basis.

It is these concerns that bring the attention of the authors to the need of an adaptive self-assessment system. The end result is the system entitled “Adaptive Self-Assessment Master” or ASAM, for short.

ASAM

ASAM is an extension to CourseMarker (CM – previously called CourseMaster), which is a computer aided assessment (CAA) system that marks programming code and diagrams and is a re-engineered version of the Ceilidh system (Benford et al, 1996). Amongst many other things, CM allows teachers to set exercises, and students to submit their solutions to these exercises on-line. Upon each submission, students obtain immediate marks and detailed feedback from the system (Higgins et al, 2002 and 2003).

ASAM is an enhancement to CM based on customising the questions to individual students by generating a personalised set of questions. With the current CM system, all students are given the same set of exercises.

In order to improve the learning environment multi-modal questionnaires that are tailored to each individual student are used. Firstly new types of question

have been added to the CM functionality, including fixed response types of question such as multiple choice and multiple answer. These are combined with currently available programming exercises and diagram based questions into a multi-modal questionnaire consisting of multiple numbers of types of question.

Question bank

Although ideas of the Item Response Theory (IRT) have been very influential in CAT systems (e.g. Abdullah and Cooley, 2000; Lilley and Barker, 2002; Syang and Dale, 1993), it is difficult to implement in practice, due to its complexity (PennState, 2000). Thus, the authors have decided to use a simpler approach to classify the questions in the question bank.

For each topic, the questions are grouped into three different levels of difficulty, namely beginner, intermediate and advanced. Teachers are advised to assign the level of difficulty to each question, corresponding to the six ascending levels of abstraction as defined in Bloom's Taxonomy (Bloom, 1956; NYIT, 2003) – the lowest three levels to 'beginner', the next two higher levels to 'intermediate' and the highest to 'advanced' (as shown in Table 1).

Level of difficulty	Corresponding cognitive aspect of learning, as defined in Bloom's Taxonomy
Beginner	Knowledge Comprehension Application
Intermediate	Analysis Synthesis
Advanced	Evaluation

Table 1 – The levels of difficulty of question items and their corresponding cognitive aspect of learning

Questionnaire skeletons

The set of questions in a given questionnaire can be of either of two modes, which is to be specified in a questionnaire skeleton. Firstly is the fixed set of questions, which are pre-selected by the teacher, from a question bank. This will mean that all students will be presented with the same set of questions. Alternatively, the teacher can opt to let the system choose a set of questions dynamically from the question bank so as to give each student a different questionnaire. With this mode, the teacher defines an adaptive questionnaire skeleton in advance, specifying the number of questions and the topic each question covers, with an option of also specifying the question type.

Self assessments

With the questionnaire skeletons in place, the system is ready for students to self-assess. When students log on to the system to attempt an adaptive self-assessment, ASAM will generate a set of questions, tailoring to each student at an appropriate level of difficulty for that student's ability.

First of all, for every question in the questionnaire, a list of questions that match the student's ability level, the topic and question type (if applicable), will be pre-selected from the question bank. It is then that ASAM chooses one from this list randomly. All the questions of each questionnaire will be displayed all in one go, and effort is made to ensure that no single question appears twice in each individual questionnaire. Nonetheless, it is still possible that the same question is presented to a student at different attempts.

Depending on the size of the question bank, it is usual that each student may be presented with different sets of questions, even when their ability level is the same at each attempt. From time to time, the system will re-adjust the ability level for each student for each topic accordingly, may it be promotion or demotion to the next level. All the profiling data, test results etc. are stored in a student model.

Figure 1 illustrates the interaction among the various components of ASAM.

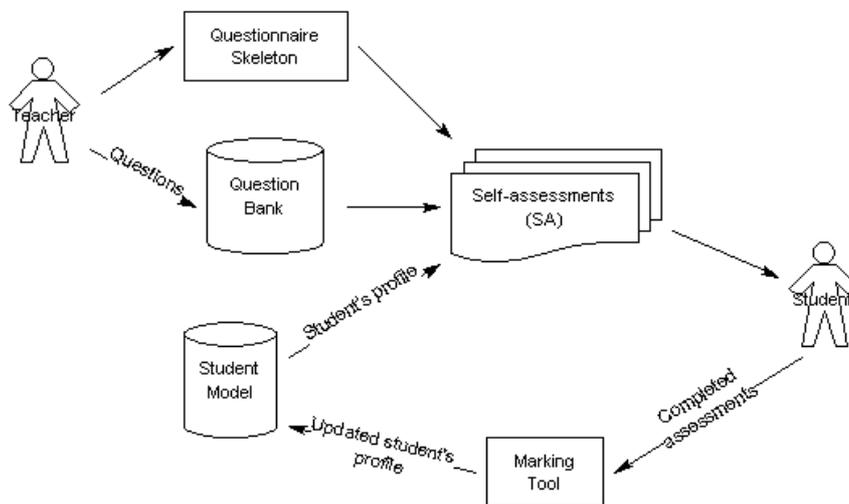


Figure 1 – The interaction among the various components of ASAM.

Using the prototype with first year students

The ASAM system has been used for the first time with the 2003/4 first-year Computer Science students for the first part of one compulsory module “Introduction to Programming” (equivalent to one sixth of the first semester work). During the first semester, they were given two self-assessments (of mainly fixed response questions) to attempt. As students were not obliged to use the prototype, not all of them used it. All in all, 89 students used the system at one point or another during the semester, whereas 85 others did not. In addition, the students were to take three multiple choice questionnaire (MCQ) tests, as part of the formal assessment. It is noted that the two self-assessments were only given to the students after the first MCQ test, with one before each of the last two tests. However, once a self-assessment is set, students can attempt it at any time during the course.

Findings

Table 2 and Figure 2 compare the results with the previous cohort of students, who had taken 3 similar MCQ tests during the similar period of time, but had no access to ASAM or any other form of self-assessment provided by the teaching staff.

Mean	N	MCQ1	MCQ2	MCQ3
2002/3	226	61.99%	45.05%	55.51%
2003/4 - ASAM	89	53.72%	44.18%	44.33%
2003/4 - no ASAM	85	50.06%	40.80%	36.41%

Table 2 – Comparison of mean results of students in 2002/3 and 2003/4

It can be observed that the scores of the 2003/4 cohort are generally lower than those of the previous cohort. Despite this, the three groups follow a similar pattern, that is, a drop in the mean score from MCQ1 to MCQ2, followed by an improvement or lesser drop for MCQ3.

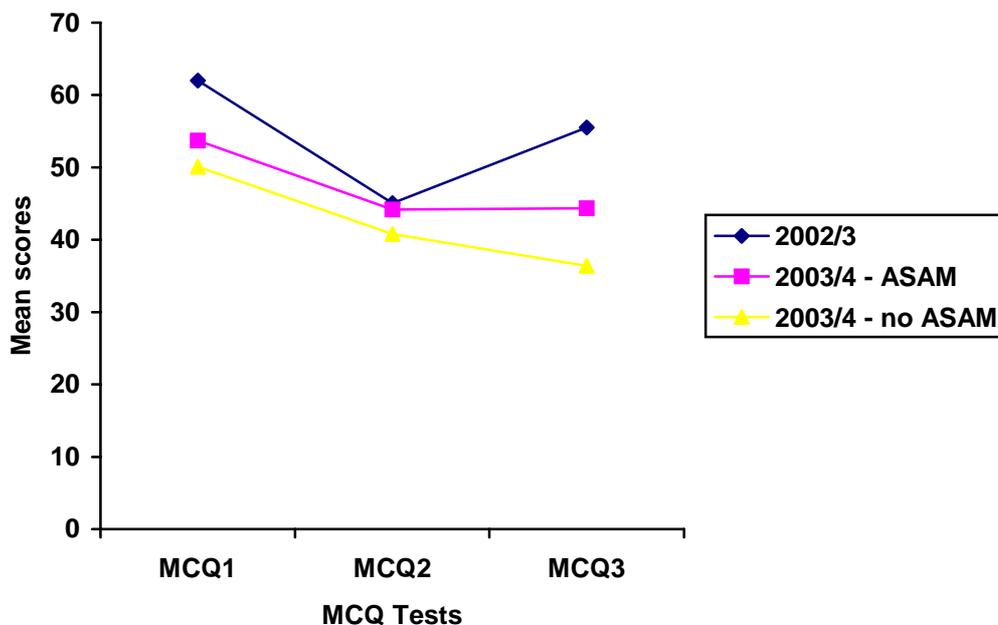


Figure 2 – Comparison of the mean score among the various groups

Focusing on the comparison between the two sets of average scores for the 2003/4 cohort, the drop in the MCQ2 scores are of similar rate. However, there was a significant improvement for the ASAM users in the MCQ3 test, whilst the non-ASAM users scored a lower average mark. Note that the self-assessment set before the MCQ2 test was of fixed mode, that is, all students were given the same set of questions at all times. On the other hand, the second self-assessment, which was set before the last MCQ test, was adaptive, that is, questions were chosen from the question bank, according to individual students' ability. From this, we can conclude that there is evidence that adaptive self-assessment is beneficial to students.

Also, the authors would like to highlight that since it was the first time that students were introduced to the ASAM system, there was some hesitation in using the system at the beginning. Upon seeing the benefits from the self-assessment after the MCQ2 test, students were then motivated to use the system more, thus leading to an improvement in the third MCQ test.

As the MCQ tests were also conducted using the ASAM system, some useful statistics from the system could also be obtained (Kleeman, 2000). This aided in identifying the weaker students, as well as the weaker topics of the whole cohort, and in turn, in preparing revision lectures.

Conclusion

From the findings, we can conclude that self-assessment is generally beneficial to students, may it be fixed or adaptive, as it allows students to evaluate their own ability informally at their own pace, thus motivating them to learn better. In addition, adaptive self-assessment adds value to their learning as they are presented with questions according to their ability level, instead of those that are either too easy or too difficult. Results from future years will give statistics that enable more rigorous assessment of the system.

It is envisioned that the system will adapt to the IMS QTI standard to store the metadata of the questions. With that in place, it will facilitate the future work of interoperating with other question banks and intelligent tutoring systems, to build up the student model, so as to enhance student's learning in an all-rounded approach.

References

Abdullah, S C, 1999, Premodelling for Examination Revision through Adaptive Testing, *3rd Human Centred Technology Postgraduate Workshop*, 30 Sep – 1 Oct 1999, Brighton.

Abdullah, S C, and Cooley, R E, 2000, Using Constraints to Develop and Deliver Adaptive Tests, *4th Computer Assisted Assessment Conference*, 21 – 22 June 2000, Loughborough.

Benford, S, Burke, E, Foxley, E, Gutteridge, N, Higgins, C and Zin A M, 1996, Software Support for Automated Assessment and Administration, *The Journal of Research on Computing in Education*.

Bloom, B, 1956, *Taxonomy of Educational Objectives*.

Eberts, R E, 1997, Computer-Based Instruction, In Helander, M, Landauer, T K and Prabhu, P (eds), *Handbook of Human-Computer Interaction*, 2nd ed., pp. 825-847.

Higgins, C, Symeonidis, P and Tsintsifas, A, 2002, The marking system of CourseMaster, In: *Proceedings of 7th Annual Conference on Integrating*

Technology in Computer Science Education (ITiCSE) 2002, Denmark, 24-26 June.

Higgins, C, Hegazy, T, Symeonidis, P and Tsintsifas, A, 2003, The CourseMaster CBA System: Improvements over Ceilidh, *Journal of Education and Information Technologies*, Vol 8, No 3, September 2003, pp. 287-304.

Kleeman, J, 2000, Getting More Meaning from the Results of CAA Discussion forum introduction, *4th Computer Assisted Assessment Conference*, 21 – 22 June 2000, Loughborough.

Lilley, M and Barker, T, 2002, The Development and Evaluation of a Computer-Adaptive Testing Application for English Language, *6th Computer Assisted Assessment Conference*, July 2002, Loughborough.

Lilley, M and Barker, T, 2003, An Evaluation of a Computer Adaptive Test in a UK University Context, *7th Computer Assisted Assessment Conference*, 8th and 9th July 2003, Loughborough.

NYIT, 2003, Bloom's Taxonomy [online]. New York: New York Institute of Technology Educational Enterprise Zone. Available at: <http://www.nyiteez.org/EDIN777/bloom.htm> [21 April 2003]

PennState, 2000, Academic Testing: Item Response Theory Approach [online]. Pennsylvania: The Pennsylvania State University, Schreyer Institute for Teaching Excellence. Available at: http://www.uts.psu.edu/Item_Response_Theory_frame.htm [17 February 2004]

Syang, A and Dale, N B, 1993, Computerized Adaptive Testing In Computer Science: Assessing Student Programming Abilities, In: *Proceedings of the 24th SIGCSE Technical Symposium on Computer Science Education*, Feb 18-19, Indianapolis, USA, pp. 53-57.

Yong, C-F and Higgins, C, 2003, Automatically Creating Personalised Exercises Based on Student Profiles, In: *Proceedings of 8th Annual Conference on Integrating Technology in Computer Science Education (ITiCSE) 2003*, Greece, 30 June-2 July.