Keing, C., Lo, J., Lam, P., & McNaught, C. (2007). Summative eAssessments: Piloting acceptability, practicality and effectiveness. In C. Montgomerie & J. Seale (Eds.), ED-MEDIA 2007 (pp. 486–495). Proceedings of the 19th annual World Conference on Educational Multimedia, Hypermedia & Telecommunications, Vancouver, Canada, 25–29 June. Chesapeake VA: Association for the Advancement of Computers in Education.

Summative eAssessments: Piloting acceptability, practicality and effectiveness

Christina Keing
Information Technology Services Centre
The Chinese University of Hong Kong
keing@cuhk.edu.hk

Judy Lo
Information Technology Services Centre
The Chinese University of Hong Kong
judyl@cuhk.edu.hk

Paul Lam
Centre for Learning Enhancement And Research
The Chinese University of Hong Kong
paul.lam@cuhk.edu.hk

Carmel McNaught
Centre for Learning Enhancement And Research
The Chinese University of Hong Kong
carmel.mcnaught@cuhk.edu.hk

Abstract: The question of whether eAssessment is suitable for summative assessment in higher education is still open. The paper reports a pilot study of summative eAssessment with 900 students in a basic English Language course. Evaluation showed that some of the concerns concerning acceptability, practicality and effectiveness are valid but most of them can be resolved through careful management, clear procedures, and suitable training for the teachers and students involved. The study found a clear acceptance of the new method by teachers and students. The operation ran smoothly. Also, the immediate feedback on students' performance effectively provided teachers with feedback on students' strengths and weaknesses. There were challenges in this pilot run – workload and discomfort in changing habits. A model of diminishing effort is proposed suggesting that the initial overheads of new technology and the tensions observed will be relieved as the system is progressively reused.

Summative eAssessments

EAssessments enable learners to interact with pre-configured questions on a computer. Technology gives eAssessment characteristics that seem to make it an approach to assessment that might outperform a paper-based counterpart under certain circumstances. EAssessment can be categorized as: 1) a strategy for learning through provision of formative assessments where students learn from interaction and feedback, and 2) a method for administration of summative assessments.

The discussion in this paper concerns those assessments which are appropriate as closed items. There is no suggestion that all, or even the majority, of assessment in higher education (or indeed at any other level of education) should be in an objective form. Obviously, assessment needs to be aligned with the desired learning outcomes (Biggs, 2003) and a variety of assessment strategies are needed in higher education programmes (Kember, & McNaught, in press, ch. 12).

Much has been written about the advantages of using technology in formative assessment. For example, Flynn, Concannon, and Campbell (2006) suggested that flexibility, variety and quality of learning are three main benefits of

eAssessment. Flexibile eAssessment enables learners to take exercises and quizzes anytime and anyplace while the questions can range from simple multiple-choice (MC) or fill-in-the-blanks items to complex context-rich and multimedia-enriched problems. Computers are excellent in handling close-ended inputs from students, while more complicated designs may allow computers to assist in teachers' judging students' performance in open-ended answers as well, though this is not common. EAssessment can provide students with additional chances to reflect on their own learning. There are many cases of formative eAssessment in the literature, which report a certain degree of success. For example, Peat & Franklin (2002) set up quizzes for a large first year biology class at The University of Sydney and found the additional exercises assisted students' learning. Ashton, Beevers, and Bull (2004) contextualized the experience of the PASS-IT project (online learning packages in Mathematics and Chemistry in Scotland) into an international context. Based on local and world-wide experiences they concluded that "rapid feedback to both right and wrong answers is a major advantage of formative computer testing" (p. 79). Flynn, Concannon, and Campbell (2006) used formative eAssessment with Accounting students at University of Limerick and found that students' performance in off-line tests were better if they had the online assessments as practice. In the local Hong Kong context, Au (2006) used case questions and a discussion board as her online assessment strategies in an art education module; she found the combined strategy to be very effective in promoting students' self-directed learning and other skills.

The use of eAssessment as a summative measure is also popular. McKenna and Bull (2000) noted that: "In the USA, approximately one million examinations for undergraduates and postgraduates were delivered and marked by computers in the 1997–1998 academic year under the auspices of a national testing programme." The drive for this popularity (even ten years ago) may be due to the increase in student numbers with relative decrease in staff resources. The new teaching and learning environment "has motivated the search for more efficient ways of assessing large students groups" (McKenna, & Bull, 2000, p. 24). EAssessment is not limited to the school setting. Reznick et al. (1992) reported a pilot examination administered by the Medical Council of Canada. The examination was part of the licencing procedures to enable newly qualified doctors and nurses to practise medicine in the country. The study affirmed the feasibility of running large-scale national examinations. Coffee, Pearce, & Nishimura (1999) discussed a major use of the technology as a staff selection and recruitment tool in the US. They studied the introduction of the online civil service examination by the Civil Service of the State of California. Results showed that the effectiveness and efficiency of the testing procedures were significantly improved by the online examination.

The practicality of summative eAssessment is more controversial than formative assessment. In the few reports in the literature concerning the processes of carrying out summative eAssessments, some authors are optimistic about the use. Encheva and Tumin (2006), for example, affirmed the feasibility in carrying out computer-aided assessments to diagnose students' progress among approximately 1000 students in a compulsory mathematic course at Stord/Haugesund University College. They noted that eAssessments can save time and effort in preparing and conducting tests, and are also efficient in supplying test results for various analytic and diagnostic purposes. These benefits seem to be more apparent when eAssessments are done on a large scale.

However, there are challenges. In a study of test mode effect, Clariana and Wallace (2002) investigated the test performance of 105 students in a freshman business undergraduate course, Computer Fundamentals. About half the students took the 100-item multiple choice test in paper-based format while the other half took its computer-based version. The results were that the computer-based tests demanded more processing load from users than paper-based tests and favoured higher-attaining students.

There are also other potential problems and challenges in running summative computer-based assessments. McKenna and Bull (2000) remarked that eAssessments require careful planning and administration. They grouped the issues that deserve close attention when developing, administering, and implementing institutional-wide computer-assisted assessments into three categories: pedagogical, managerial, operational. Zakrzewski and Steven (2003) also highlighted the importance of quality control in large-scale eAssessments. MacKenzie (2003) warned that the new method may not be well perceived by teachers and students as it is very different from usual practice. There are also concerns related to technical difficulties, security of the system and fairness of treatment. All these factors may decrease the enthusiasm of teachers and students for the strategy.

In view of the growing popularity of using eAssessments, more research into the characteristics of the approach is needed. McKenna and Bull (2000) suggested that the stakes are high in running summative eAssessments as they

often involve a large amount of preparation time, and a large number of students, teachers and supporting personnel. Quality has to be ensured and this requires clear understanding of the strategy.

Context of the Study

The Chinese University of Hong Kong (CUHK) is a traditional, comprehensive, research-intensive university. The use of eLearning is very much as an adjunct strategy and is not high overall, even for formative assessment. A study of the University's eLearning in 2004 (McNaught, Lam, Keing, & Cheng, 2006) showed a very low use of the quiz function of WebCT (Figure 1). The team for this study knew that, as eAssessment is not common at CUHK, it was important that we pay particular attention to the design, implementation and evaluation of the pilot study on summative eAssessment. The evidence from this study is likely to significantly influence future work in the area of eAssessment at CUHK.

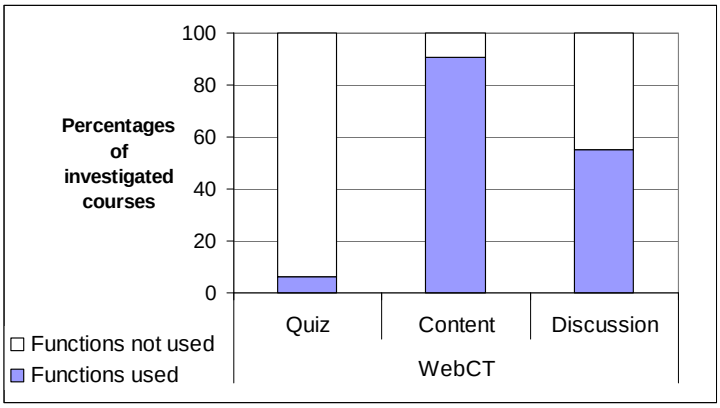

Figure 1: Web functions in the websites hosted in WebCT (2004 data)

Objectives of the Study

The paper reports a pilot study at CUHK to evaluate the practicality of eAssessment as a form of university-wide summative assessment. Our starting point was that we considered that eAssessment could be a very useful innovative assessment strategy if we took enough precautions to avoid the potential problems, and used it with appropriate types of assessments. EAssessment systems can randomly pick question items from question banks to easily generate separate tests for students in different test sessions. EAssessments can also extract, compile and analyze students' answers to questions with relative ease. These aspects make eAssessments potentially especially useful in courses or programmes with a large number of students and for tests that tend to be run regularly because the benefits balance the preparatory effort of setting up the system and question banks. EAssessments are likely to be most useful for tests that concern factual knowledge and basic concepts as they are relatively weak in handling open-ended questions. Thus, they may be more suitable with introductory courses in which closed-ended items such as MC questions are often the preferred format of testing.

In this study eAssessment was conducted with about 900 students in a basic English Language course.

Rather than adopting a strict research approach, our study tended towards a naturalistic model (Guba & Lincoln, 1981; Alexander & Hedberg, 1994). Through observing the practicality of the new strategy in a realistic setting, we expected that the lessons we learnt would be readily transferable to other eAssessment projects. Both qualitative and quantitative data were collected. We had three research questions for the trial:

1. Acceptance: Did both teachers and students feel comfortable with the new method?
2. Practicality: Was the administration and implementation of eAssessment smooth and trouble-free?
3. Effectiveness: was the eAssessment an effective tool for the teachers to understand their students' performance?

Three groups at the University were involved in the study: the English Learning and Teaching Unit (ELTU), the Information Technology Services Centre (ITSC) and the Centre for Learning Enhancement And Research (CLEAR). One teacher and one administrative staff from ELTU were responsible for coordinating with the other teachers in designing and conducting the assessment. Two computer officers from ITSC assisted in providing consultation and developing the IT solutions. Two researchers from CLEAR assisted in tracking and reviewing the pilot study.

In 2006, the ELTU designed a new language programme called the Common English Foundation (CEF) which was to be incorporated alongside Level 1 ELT Courses for first-year students. An extra hour was added to these courses and a language enhancement programme consisting of grammar and vocabulary was taught. This programme comprised 900 students in 45 classes taught by 21 ELTU teachers.

Towards the end of the semester, all students sat an eAssessment which contributed 20% towards the final course grade. The duration of the eAssessment was one hour and consisted of two sections: Section 1 contained 40 MC questions to test students on the grammar content of the course, and Section 2 contained another 40 short-answer questions to test students on vocabulary, including verb use, phrasal verbs, extended vocabulary and word formation.

Since the CEF course involved a large group of students, teachers had the following concerns regarding the assessment:

1. There would be a heavy workload in marking paper examination papers.
2. Since it is not possible for 45 classes to take the examination at the same time, classes were scheduled to take the examination at different time slots within one week.
3. Security was a concern in that teachers didn't want the examination questions to be disclosed easily and wanted students to work on different, but equivalent, examination papers.
4. Since this course and the assessment were both new, teachers wanted to analyze the performance of students on each question and make necessary adjustments to the question items in subsequent administration of the tests.

To address these concerns, CEF teachers decided that eAssessments might be a sensible solution rather than a traditional paper-based version. ITSC advised that the system could be built using the online quiz function of the university-supported eLearning platform, WebCT. This would be beneficial mostly because:

1. For MC and short-answer questions, the platform supported auto-marking which would significantly save teachers' time.
2. It was possible to randomly select questions from a question bank and compile a different examination paper for each student even during the same examination session.
3. Since all answers were saved electronically into the database, it would be very convenient for teachers to do item analysis and performance analysis. Teacher should be able to get a summary of all submitted answers and the corresponding frequency for revision and necessary changes for future examinations.

The examinations were conducted in four computer labs. The 45 classes were assigned to different examination time slots within one week. During the peak period, there were about 150 students working on the examinations at the same time.

In order to evaluate the practicality of eAssessment, data were collected from three groups of stakeholders in the pilot study: ELTU teachers, ITSC staff, and students. We held individual interviews with the coordinators from the ELTU. We have also met with a group of five other ELTU teachers who assisted with the computer lab administration during the examinations. Data from the two ITSC computer staff members were collected from their informal conversations with the researchers. They also constantly referred to the logs and email communications they held with all the other staff involved in the project as a means to consolidate their thoughts. Lastly, data from the students were collected through an online survey administered after the exam. All students in the cohort were invited to take the survey. 183 replies were collected and the response rate was approximately 11%.

Precautions Taken in the Study

In compliance with the warnings of McKenna and Bull (2000) that attention should be paid to pedagogical, managerial and operational issues of eAssessments, various measures were taken to ensure quality in all these areas.

Pedagogically, we wanted to ensure that the tests tapped appropriate learning levels with regard to the learning objectives of the course. We imposed measures such as:

- Care paid to the writing of individual question items: Questions collected from all ELTU teachers were carefully reviewed, selected and categorized before being added to the CEF question database on WebCT. Care was taken to ensure that the questions were of similar difficulty so that the system could generate tests of reasonably equivalent levels for all students.
- Diversifying question types: The CEF course panel committee decided to use MC questions and short-answer-type questions in the test. These two types of questions allowed auto-grading to be performed by the system. Two formats were used to maintain students' interest and motivation during the test. Also, as students were required to actually type answers in the short-answer-type questions, students were tested on their ability with language 'production', which was in line with the basis of CEF's objective of improving students' writing.
- Compilation of a question bank and a sound question selection mechanism: More than 650 question items were generated and put into the question bank. These questions were categorized by topics. Selection was made of a certain number of questions from each topic so that all individual tests were composed of a balanced number of questions across all topic areas.

From a management perspective, we wanted to make sure that teachers and students were all trained in using the new method and were provided with clear instructions and guidelines. We employed strategies such as:

- Training staff to equip them with appropriate new skills: In order to get all the teachers become familiar with the test system, they were encouraged to attempt both the practice test and the real test together with the students. The committee regularly announced updates on the test through email to all teachers. A total of three newsletters, together with instructions to invigilators, were prepared and distributed.
- Training students: In order to get all of the test-takers become familiar with the test system, a practice test was set up for test-takers to trial before the real test. Class teachers could also opt to take their class to a test centre to let the class trial the practice test together under supervision.
- Writing up examination regulations and test protocols: In order to provide all test-takers a fair ground to start working on the real test, an instruction sheet was prepared for test-takers. Invigilators would brief the test-takers on all the important notes on the system following the instruction sheet before begin of the real test.

Operationally, strategies were employed to ensure smooth running of the tests. Concerns and the solutions adopted included:

- Clear timetabling: There were altogether 17 test sessions and up to five classes might conduct the test at different remote and scattered test centres around the campus at the same time. A dedicated central coordinator whom all teachers could reach by phone at all times was appointed. This central coordinator worked together with ITSC computer staff to ensure smooth operations during the test period.
- Checking of equipment: Test centres were equipped with hardware of different standards. The team conducted site visits to ensure that all computers met with the minimum requirement and that the differences between different computers would not greatly impact the performance of test-takers using the computers.
- Setting contingency arrangements: Strategies to handle various contingency situations were planned beforehand. For example, if a test-taker reported that something was wrong with a question during an actual test session, invigilators would record the student id and the question id of the problematic question. The question would be frozen and would not be used in the following sessions. Some computers might fail during the test session. As the test-taker had not submitted the test for grading, he/she could continue to work on the test using a spare computer in the test centre. The case would be reported to the central coordinator and the panel committee would decide whether extra marks should be given to compensate the test-takers for the loss of time during the computer break-down on a case by case basis.
- Ensuring security and fairness: Special accounts were created for test-takers to use during the test periods. These accounts were distributed to test-takers at the test centres just before the test began. Security could thus be assured and unauthorized access to the test system could thus be avoided. Any interruption might cause serious impact on test-takers' performance and affect the fairness of the test. Several measures were taken to ensure a safe and stable test environment. On the system level, load tests were conducted to confirm that the system was capable of handling the load generated during the test period (about 150 concurrent users as a maximum).

- Countering cheating: Test centres were not arranged and furnished solely for test purposes and thus there was no partition between computers at the test centres. Test-takers would be able to peek at others' computer screens. Test items were thus configured to be display in a random sequence for each of the test-takers and they were randomly selected from the question bank. That each test-taker who sat the test received a different permutation of questions; this helped to ensure fairness of the test.

Evaluation

In this section the data received from the various evaluation strategies are discussed. Both advantages and weaknesses were found concerning the use of the technology in the test sessions. However, we were cognizant that this was the time first the test was run in this manner. Some of the challenges encountered may not be real long-time problems of eAssessment but relate to temporary hardships in the transitional phase.

Acceptability

The teachers seemed to accept the new method. In general, they found the new technology easy to manage. The training sessions went well. Teachers needed to do basic operations with the WeCT quiz function. Although the ELTU did not participate in the actual uploading of the quiz items and the creation of quiz logics in the pilot test, the Unit thought they would take over future maintenance of the question bank and the testing system from ITSC starting in the next round. The Unit felt confident about further modifying and enlarging the question bank. They also felt they could create test sessions, assign students to sessions, and host these sessions with relative ease.

The one thing most commonly mentioned by the teachers as a negative factor was the added workload. Though the ELTU teachers in general remarked that, as in pen-and-paper tests, much time is needed in designing and collecting questions from all teachers, much more time now was required in the online version to prepare for the randomization needed for the online system. The questions had to be carefully reviewed, selected and categorized by topic before input into the question database. After removal of extra difficult and long questions, 376 MC questions were categorized into 26 categories, and 278 short-answer-type questions were categorized into 30 categories to form different question sets. Each of the test sets contains 80 question items of similar difficulty level to ensure fairness of the test. The categorization and the reviewing took a great deal of time. Also, it took longer than expected to design the actual question selection processes: e.g. how many categories to have, how many questions to select in each category, how many test sessions to hold, and how many questions could be similar from one test to another, etc.

Not all ELTU teachers were IT-competent. In order to prepare them for the test, the coordinators had to prepare newsletters and emails to keep all the teachers updated with the development of the test. Detailed instructions were also provided to invigilators to make sure that all of them understood the progress and the workflow on test days.

The ITSC computer staff also had a heavy workload in importing the questions into WebCT question database, setting up the test on WebCT for conditional release during the test period, and preparing analysis reports after the tests.

As for the students, their acceptance of the new method was also not unanimous. The responses to the survey question about "I prefer the online test to a pen-and-paper test" are represented in Figure 2. In general, more students preferred eAssessment to pen-and-paper but there are still a significant group of students who strongly favoured the traditional mode.

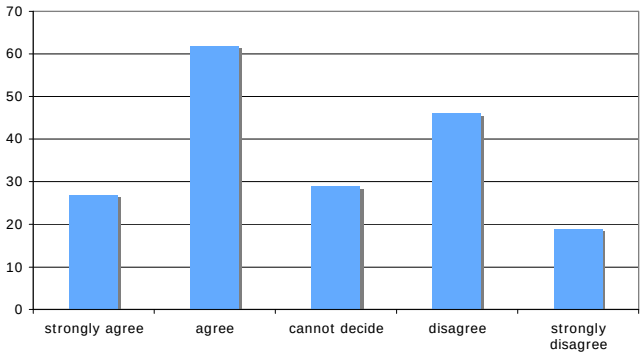

Figure 2: Students' preference for online tests over pen-and-paper tests

The open-ended questions on the survey also showed the same dichotomy of views. On the one hand there were very positive remarks concerning the new method such as (slightly edited for ease of reading):

- "It is more convenient than a pen-and-paper test."
- "Not so tiring as a pen-and-paper test."
- "Not as nervous as doing a paper test."
- "Easy to see which question I have answered."
- "I can amend my answer without using a correction pen."
- "The timers on computers are accurate."

On the other hand, there are also negative remarks concerning eAssessment:

- "It is better to use the traditional pen-and-paper test."
- "Too tiring for eyes."
- "I can't collect my thoughts in front of a computer."
- "The quality of the mouse and the computer has an effect on performance."
- "It would be unfair for those who type words slowly and wrongly."
- "It is not fair that the questions are not the same for all students. To make the test fair, all students should answer the same set of questions."

Practicality

The actual running of eAssessment was smooth and quite error-free. There were no reports from either ELTU teachers or ITSC staff about major operational problems. Students also commented favourably on the procedures. They responses to the question "The online test was conducted smoothly" are represented in Figure 3. As illustrated, the vast majority of the students agreed to the statement.

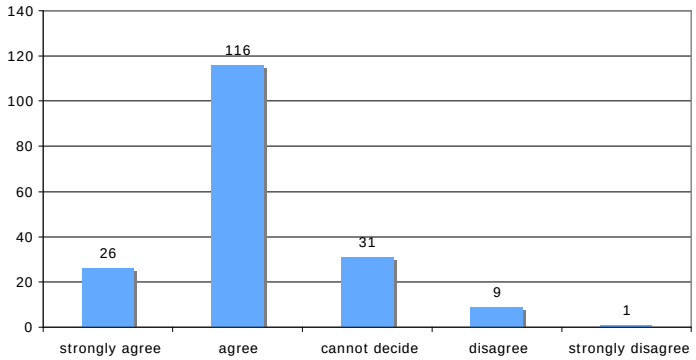

Figure 3: Students' perceptions of the smooth conduct of the online tests

Effectiveness

The reporting of students' performance was a more complicated task than expected. It took ITSC about one day to import all the questions selected to be used in the test into the question database. After that, a half-day's work was required to match each of the scores with the randomized question items each student took and to group the test scores according to the test sessions the students were in. All the test data (students' choice for test items) were imported into a database for further processing. Reports that could facilitate detailed item analyses were also prepared using the business analysis tools bundled with the database management software. It took ITSC an additional half-day to generate these reports.

The reporting function of eAssessment was indeed very useful. All of the ELTU teachers agreed that the auto-grading feature saved individual efforts on manual grading. The immediate feedback from the system on students' performance effectively provided teachers with feedback on students' strengths and weaknesses. For example, teachers obtained information about the areas students were weak in by viewing the score distribution of the reports. Teachers' effort can then be put into making sense out of the students' scores and converting them into remedial actions for future directions in teaching, rather than spending the time on getting the marks ready only. Item analysis reports, which might not be readily available for pen-and-paper tests, can also be prepared easily and in a timely manner. If the test had been administered using traditional pen-and-paper format, some reports (such as the reports for item analyses) might not have been technically or economically possible.

Discussion

Benefits

Many of the expected strengths of eAssessment were confirmed in the pilot study. EAssessments can randomly pick question items from question banks to easily generate separate tests for students in different test sessions. EAssessments can also extract, compile and analyze students' answers to questions with relative ease. These make eAssessments useful in assessing students' performance especially in courses or programmes with a large number of students and for tests that tend to be run regularly because there can be a substantial return on the investment in setting up the system and question banks in the first place.

The study found a considerable acceptance of the new method by both teachers and students. The arrangements and operations were by and large smooth and successful. After the pilot study, teachers at ELTU decided to continue to use eAssessment for the CEF course and will continue to add questions to the question bank in WebCT to improve the randomization and variety of the exam. They also appreciated the comparatively speedy and care-free functions of the auto-grading function of the testing system.

Problem of the large initial overhead

Workload was an important negative factor recorded in the study. Both the teachers and the ITSC staff reported that a great deal of time had been spent on the preparation and administration of the tests. However, many of the causes of the workload are 'one-off' in the sense that they are no longer required (or less effort will be required) when the system is reused in the future rounds. For example, the questions in the pool can be reused. The training guidelines and materials can be the same. Teachers do not need to be trained again; only new colleagues will need training. The mechanism to automatically select question items to form separate tests is programmed into WebCT and can stay unless radical changes are needed. Thus, we tend to regard that the workload, despite being quite heavy to both teachers and the supporting team, is an initial overhead of eAssessments, and should be resolved over time.

Problem of changing ingrained habits

There were clearly some obstacles related to usability. Quite a large proportion of students in our study were not completely satisfied with the eAssessment. Some of their negative remarks seem to be related to inherit characteristics of the technology (such as the difficulty of reading from the screen); many, however, are related to the skills and habits of students such as the ability to type, and how effectively they can 'think' in the computer lab

environment. These habitual factors can change over time. With more training on using the technology for the next cohort of students, and a general increase in students' overall IT skills, resistance from the students should decrease.

We, therefore, regard that many of the challenges met in this pilot study are threshold obstacles and are significant only in the first, or first few, attempts when a great deal of background work is needed and teachers and students are still in a transitional period. The situation is depicted in Figure 4 as a model of diminishing effort. Workload and discomfort in changing of habits are the initial overhead of the new technology. The tension observed in the pilot test will be relieved when the system is reused and students are more and more computer-ready. The benefits from eAssessment will outrun the effort required in the long run.

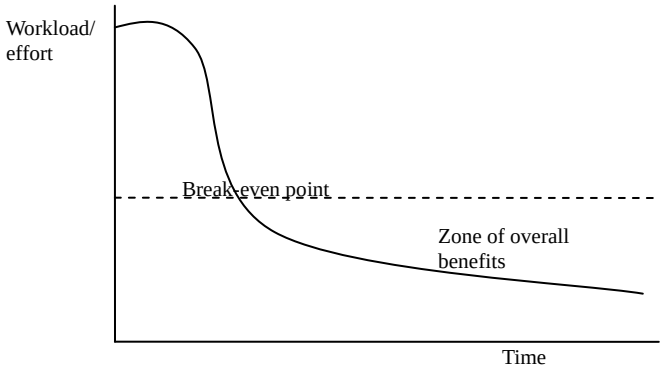

Figure 4: Model of diminishing effort for eAssessment

Another lesson learnt from this pilot test is that good planning is definitely needed in implementing eAssessment in order to accommodate the high overhead. The planning is not limited to the logistics and strategies to handle contingency situations, but also focuses on the determination to target it as a sustainable model of testing from the very beginning.

Conclusion

The pilot study was our first explorations into the acceptance, practicality and effectiveness issues concerning the use of eAssessment as a large-scale testing tool. The preliminary results were on the whole positive although the acceptability of the method could be further improved. Evaluation of the pilot study showed that some of the common concerns concerning acceptability, practicality and effectiveness are valid but most of them can be resolved through careful management, clear procedures and suitable training to the teachers and students concerned. In particular, there were challenges concerning workload of teachers and the change of habits imposed on students; and these need to be addressed in the forthcoming trials of the method. More about the strategies will be learned in our future attempts to extend and refine the strategy based on our experience thus far.

The present trial is also limited in the sense that it did not study the potential strengths of eAssessments in handling multimedia materials (e.g. Tschirner, Muller, Pfeiffer & Thomsen, 2006). The research team would very much like to investigate this aspect in our future work.

Acknowledgement

The commitment and expertise of staff in the English Language Teaching Unit at CUHK is gratefully acknowledged. Special thanks and appreciation go to the teaching assistant and all the ELTU staff involved with the CEF course.

References

Alexander, S., & Hedberg, J. (1994). Evaluating technology-based learning: Which model? In K. Beattie, C. McNaught & S. Wills (Eds.). *Multimedia in higher education: Designing for change in teaching and learning*, (pp. 233–244). Amsterdam: Elsevier.

Au, E. (2006). Using online processes to enhance assessment and learning. In D. Carless, G. Joughin, N. F. Liu, & Associates (Eds.). *How assessment support learning: Learning-oriented assessment in action* (pp. 52–65). Hong Kong: Hong Kong University Press.

Ashton, H. S., Beevers, C. E., Bull, J. (2004). Piloting e-assessment in Scottish schools – Building on past experience. *International Journal on E-Learning, 3*(2), 74–84.

Biggs, J. B. (2003). *Teaching for quality learning at university* (2nd ed.). Buckingham: Society for Research into Higher Education & Open University Press.

Clariana, R., & Wallace, P. (2002). Paper-based versus computer-based assessment: Key factors associated with the test mode effect. *British Journal of Educational Technology, 33*(5), 593–602.

Coffee, K., Pearce, J., & Nishimura, R. (1999). State of California: Civil service testing moves into cyberspace. *Public Personnel Management, 28*, 283–300.

Encheva, S., & Turmin, S. (2006). Web-based assessment system for teaching mathematics on undergraduate level. *Advanced Technology for Learning, 3*(1), 9–14.

Flynn, A., Concannon, F., & Campbell, M. (2006). An evaluation of undergraduate students' online assessment performances. *Advanced Technology for Learning, 3*(1), 15–21.

Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation*. San Francisco: Jossey-Bass Publishers.

Kember, D., & McNaught, C. (in press). *Enhancing university teaching: Lessons from research into award-winning teachers*. London: Routledge Falmer.

MacKenzie, D. (2003). Assessment for e-learning: What are the features of an ideal e-assessment system? Proceedings of the 7th CAA Conference, 8–9 July. Retrieved November 21, 2006, from http://magpie.lboro.ac.uk/dspace/bitstream/2134/1914/1/mackenzie03.pdf

McKenna, C., & Bull, J. (2000). Quality assurance of computer-assisted assessment: Practical and strategic issues. *Quality Assurance in Education*, 8(1), 24–32. Retrieved November 21, 2006, from http://www.emeraldinsight.com/Insight/ViewContentServlet?Filename=Published/EmeraldFullTextArticle/Articles/1200080103.html

McNaught, C., Lam, P., Keing, C., & Cheng, K. F. (2006). Improving eLearning support and infrastructure: An evidence-based approach. In J. O'Donoghue (Ed.). *Technology supported learning and teaching: A staff perspective* (pp. 70–89). Hershey, PA: Information Science Publishing.

Peat, M., & Franklin, S. (2002). Supporting student learning: The use of computer-based formative assessment modules. *British Journal of Educational Technology, 33*(5), 515–523.

Reznick, R., Smee, S., Rothman, A., Chalmers, A., Swanson, D., Dufresne, L., Lacombe, G., Baumber, J., Poldre, P., & Levasseur, L. (1992). An objective structured clinical examination for the licentiate: Report of the pilot project of the Medical Council of Canada. *Academic Medicine, 67*(8), 487–494.

Tschirner, N., Muller, M., Pfeiffer, O., & Thomsen, C. (2006). Design and realization of multimedia-examinations for large numbers of participants in university education. *International Journal of Emerging Technologies in Learning, 1*(2), 1–4.

Zakrzewski, S., & Steven, C. (2003). Computer-based assessment: Quality assurance issues, the hub of the wheel. *Assessment & Evaluation in Higher Education, 28*(6), 609–623.