



Higher Education Research and Development Society of Australasia, Inc

# Learning for an Unknown Future

*Proceedings of the*

## **26<sup>th</sup> HERDSA Annual Conference**

6-9 July 2003

Christchurch, New Zealand

Gemmiti, F. (2003) Did Human Biology students use, recommend and benefit from computer-based assessment?, in *Learning for an Unknown Future, Proceedings of the 26th HERDSA Annual Conference, Christchurch, New Zealand, 6-9 July 2003: pp 231.*

Published 2003 by the  
Higher Education Research and Development Society of Australasia, Inc  
PO Box 27, Milperra, NSW 2214, Australia  
[www.herdsa.org.au](http://www.herdsa.org.au)

ISSN: 0155-6223  
ISBN: 0 90 8557 55 8

This research paper was reviewed using a double blind peer review process that meets DEEWR requirements. Two reviewers were appointed on the basis of their independence, expertise and experience and received the full paper devoid of the authors' names and institutions in order to ensure objectivity and anonymity. Where substantial differences existed between the two reviewers, a third reviewer was appointed. Papers were evaluated on the basis of originality, quality of academic merit, relevance to the conference theme and the standard of writing/presentation. Following review, this full paper was presented at the international conference.

Copyright© 2003 HERDSA and the authors. Apart from any fair dealing for the purposes of research or private study, criticism or review, as permitted under the Copyright, Design and Patent Act, 2005, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms and licenses issued by the copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers at the address above.

# Did Human Biology students use, recommend and benefit from computer-based assessment?

**Frank Gemmiti**

University of New South Wales, Australia  
fgemmiti@anhb.uwa.edu.au

**Abstract:** *Computer-based formative and summative assessment is becoming more widespread in higher education. However, implementation of computer-based assessment must be followed by an evaluation of staff intentions. In 2002, staff of a dual-semester, first-year, Human Biology unit at The University of Western Australia first implemented computer-based practice quizzes and graded tests. Evaluation of the computer-based assessment system involved two methods. First, a student questionnaire administered in semester two elicited information about awareness, use and opinion of the semester-one practice quizzes and graded tests. Next, statistical investigations were conducted into whether increased use of practice quizzes improved graded-test performance and whether increased use of practice quizzes and graded tests increased general-unit performance. All respondents were aware of the practice quizzes and graded tests and almost all respondents used and recommended the practice quizzes and graded tests. Some respondents suggested minor improvements to the computer-based system. Increased use of practice quizzes was associated with better performance on graded tests, despite general-unit performance. Although an association existed between general-unit performance and use of practice quizzes and graded tests, the nature of the association was inconclusive. Further analysis is required to confirm the association between use of practice quizzes and graded-test performance and the nature of the association between general-unit performance and use of practice quizzes and graded tests.*

**Keywords:** *computer-based assessment; formative assessment; summative assessment*

## Introduction

Increasing use of computer-based assessment in higher education reflects increasing student-staff ratios and the potential of advancing technology to extend assessment methods (Thelwall, 2000). Computer-based methods have been used for formative assessment to promote learning and for summative assessment to grade performance (Dalziel, 2001). There are many possible advantages of using computer software for formative and summative assessment (Charman, 1999). Through automatic generation of multiple versions of a test on the same subject, students can undergo repeated practice quizzes or can take open-access graded tests with little opportunity to copy answers from others. Immediate feedback to students enables the alteration of an activity once performed, thereby preserving the link between action and reaction. Results become available to staff once students have logged out, making administration efficient. The removal of biases in staff judgement improves equity of

assessment. Individual learning styles can be accommodated by various computer facilities, such as graphics, sound or animation. Independent learning is encouraged by the flexibility of online access.

Cross-disciplinary evidence suggests that students favour computer-based assessment (e.g., Bull and Hesketh, 2001; Dalziel and Gazzard, 1999; Proctor and Donoghue, 1996; Sambell, Sambell & Sexton, 1999), even where computers are not a regular component of learning (Pritchett & Zakrzewski, 1996) and increased use of computer-based practice quizzes improves performance on computer-based graded tests (Sly & Rennie, 1999; Sly & Stace, 1999), despite general ability (Sly & Western, 1998). There has been concern about the capacity of simple, computer-based tests to foster higher-order thinking (Laurillard, 1993, p. 153), the mismatch between computer-based continuous assessment and pencil-to-paper final examinations, unequal familiarity with computers among students, technical difficulties with a computer system, the degree of feedback provided by a programme and the opportunity for students to confer on answers when taking an open-access graded test (Harvey & Mogyey, 1999). True validity and reliability of computer-based assessment are difficult to evaluate because establishing what is assessed (validity) and how consistent the assessment (reliability) is requires many subjects, investigators and settings (Perkin, 1999). However, given the possible advantages and disadvantages of computer-based assessment, the intended effects on students must be evaluated as soon as possible (Bull, 1994).

### ***Computer-based assessment in a first-year Human Biology unit***

In 2002, staff of a dual-semester, first-year Human Biology unit at The University of Western Australia first implemented computer-based, formative and summative assessments. A practice quiz and a graded test were available on computer for each of six topics during each semester. Staff intentions for the practice quizzes and graded tests were as follows:

- 1) To have students undertake continuous assessment of basic knowledge
- 2) To provide students with a positive experience of continuous assessment
- 3) To have practice quizzes improve graded-test performance
- 4) To have practice quizzes and graded tests improve general-unit performance

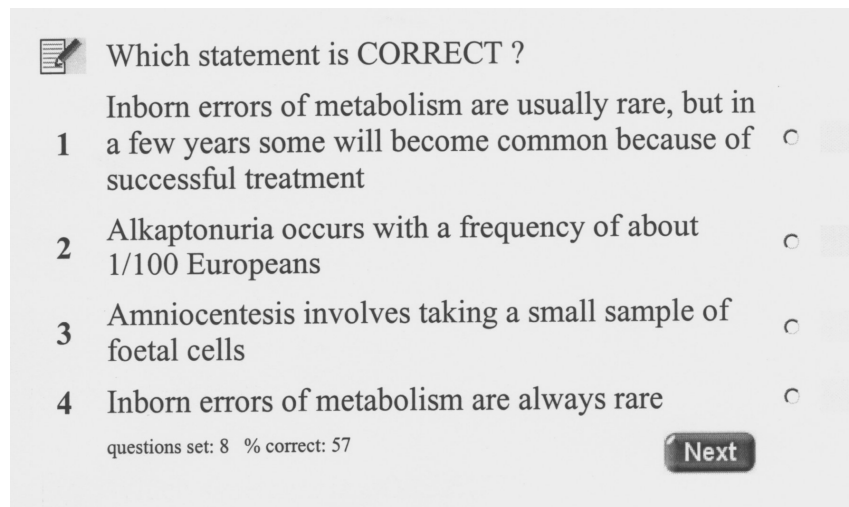
The author of this paper had an opportunity to evaluate the practice quizzes and graded tests at a time when students were accessible. This paper details the evaluation of the practice quizzes and graded tests implemented in semester one of 2002, according to staff intentions.

## **Materials and methods of evaluation**

### ***Practice quizzes and graded tests***

The practice quizzes and graded tests were composed of multiple-choice questions. A question was set when four knowledge statements became combined under a stem of “Which statement is CORRECT?”, “Which statement is INCORRECT?” or “Identify the structure labelled A”. Only one of four knowledge statements was the answer to a stem. A student was required to click on a radio button beside the chosen statement (see Figure 1). The six topics in order of availability were “evolution”, “cells, cell division & protein synthesis”, “genetics”, “tissues, skin & anatomy”, “reproduction and embryology” and “hominid evolution and primates”. Questions for each topic were composed from random permutations of hundreds of relevant knowledge statements. Students could practise unlimited questions without a time restriction for each practice quiz but could answer only 15 questions within 15 minutes for each graded test. Questions for each graded test were composed from the same bank of knowledge statements as the corresponding practice quiz. An answer to a practice-quiz

question prompted feedback in the form of “right” or “wrong” unlike an answer to a graded-test question. An answer to a graded-test question was immutable once selected. Students received a percentage grade once



The screenshot shows a question interface with a pencil icon and the text "Which statement is CORRECT ?". There are four numbered options, each with a radio button to its right. At the bottom left, it says "questions set: 8 % correct: 57". At the bottom right, there is a "Next" button.

Which statement is CORRECT ?

- 1 Inborn errors of metabolism are usually rare, but in a few years some will become common because of successful treatment
- 2 Alkaptonuria occurs with a frequency of about 1/100 Europeans
- 3 Amniocentesis involves taking a small sample of foetal cells
- 4 Inborn errors of metabolism are always rare

questions set: 8 % correct: 57

Next

**Figure 1. A computer-based multiple-choice question**

all 15 questions to a graded test were answered or once 15 minutes expired. All graded tests contributed five percent to the annual grade of a student. Graded tests were available for a finite period each semester: Closing dates occurred at least two weeks after the relevant lectures and tutorial. The practice quizzes and graded tests could be attempted only from a computer terminal in a video-monitored computer room on campus.

### ***Subjects and data collection***

During tutorial classes in semester two, 2002, students were given a questionnaire designed to elicit information about awareness, use and opinion of the semester-one practice quizzes and graded tests. Each student was requested to provide a personal ID number on the questionnaire so that responses could be linked to grades for statistical analysis. A space was provided in the questionnaire for comments. All comments were paraphrased by the author of this paper and then examined for the capacity to be arranged into categories agreed upon by the author and two staff members.

### ***Statistical analysis***

To investigate whether use of practice quizzes improved graded-test performance, Pearson's correlations were conducted on average test grades compared with the number of practice quizzes attempted and total time spent on practice quizzes. Independent relationships with average test grade were also examined. Since the average test grade may have varied with general-unit performance or familiarity with the conditions of graded tests, semester-one unit grade and number of graded tests attempted were included together in a multivariate analysis with the number of practice quizzes attempted and the total time spent on practice quizzes. The contribution of all graded tests to the semester-one unit grade was removed to avoid double counting. To investigate whether more time spent on a practice quiz improved performance on the corresponding graded test, ANOVAs were conducted for each topic using Bonferroni, *post hoc* analyses. To investigate whether use of practice quizzes and graded tests improved general-unit performance, Pearson's correlations were conducted on the semester-one unit grade compared with the average test grade, number of graded tests attempted,

number of practice quizzes attempted and total time spent on practice quizzes. As with the average test grade, independent relationships with the semester-one unit grade were also examined. All variables were investigated for relationships with the semester-one unit grade using a Pearson correlation and included together in a multivariate analysis. Pearson's correlations were two-tailed since the study was exploratory. Percentage variables were arcsine transformed and multiple linear regression was used for multivariate analyses since all independent variables were continuous (Sokal & Rohlf, 1981, pp. 158, 427-28, 459-60). Only questionnaire respondents were included in statistical analyses.

## Results

### ***Student awareness and use of practice quizzes and graded tests***

All questionnaire respondents (381) were aware of the practice quizzes and graded tests. Respondents became aware in two main ways: at a lecture (40% of 380 respondents) and at a tutorial (36.6%). Other respondents became aware via the tutorial manual (15.5%), a friend (4.5%), the unit web site (3.2%) or the unit notice board (0.3%). Most respondents attempted all practice quizzes (81% of 379 respondents). The most common reason that respondents cited for missing any practice quiz was "no time" (64.8% of 125 reasons); other reasons were "probably unhelpful" (8.8%), "computer trouble" (7.2%), "inconvenience" (6.4%), "unaware" (5.6%), "disinclination" (3.2%), "forgot" (2.4%) and "incapacity" (0.8%). Most of the 508 enrolled students attempted all graded tests (68.3%). The most common reason that respondents cited for missing any graded test was, as with practice quizzes, "no time" (47.9% of 119 reasons); other reasons were "forgot" (13.4%), "inconvenience" (10.9%), "unaware" (7.6%), "computer trouble" (7.6%), "personal mistake" (5.0%), "incapacity" (2.5%) and "disinclination" (1.7%).

### ***Student opinion of practice quizzes and graded tests***

Almost all respondents regarded the practice quizzes as "very useful" (78.7% of 380 respondents) or "useful" (17.6%) for the graded tests. Most respondents regarded the practice quizzes and graded tests as "useful" (43.4% of 380 respondents) or "very useful" (34.7%) for the final exams. Almost all respondents would give a recommendation for the practice quizzes and graded tests to future students in the unit (94.5% of 380 respondents).

Comments covered five main areas. There were comments on the computer interface for the practice quizzes and graded tests, such as the design of questions ("Format"). There were comments on how the practice quizzes and graded tests were run in the unit, such as the suitability of graded-test closing dates ("Operation"). There were comments on the perceived learning benefit from attempting the practice quizzes and graded tests, such as the directive force on study ("Learning Value"). There were requests for extensions to the practice quizzes and graded tests, such as a mixed-topic practice quiz and graded test ("Extension"). There were comments about the existence of the practice quizzes and graded tests ("Philosophy"). Table 1 shows the five areas and associated comments.

### ***Did use of practice quizzes improve graded-test performance?***

Average test grade increased as more practice quizzes were attempted ( $r = 0.426$ ,  $p < 0.001$ ,  $n = 379$ ) and as more time was spent on practice quizzes overall ( $r = 0.568$ ,  $p < 0.001$ ,  $n = 362$ ). Independent correlates of average test grade were (1) semester-one unit grade ( $B = 0.802$ ,  $p < 0.001$ ), (2) number of graded tests attempted ( $B = 0.129$ ,  $p < 0.001$ ) and (3) total time spent on practice quizzes ( $B = 0.035$ ,  $p < 0.001$ ), having explained about 80 percent of the variance.

As more time was spent on a practice quiz for all topics, performance on the corresponding graded test increased. All ANOVAs were significant beyond the one-percent level. *Post hoc* analyses revealed that for all topics, respondents who spent more than one hour on the practice quiz gained a higher grade on the corresponding graded test than respondents who spent less than 40 minutes. Figure 2 shows ANOVA results for the “evolution” topic.

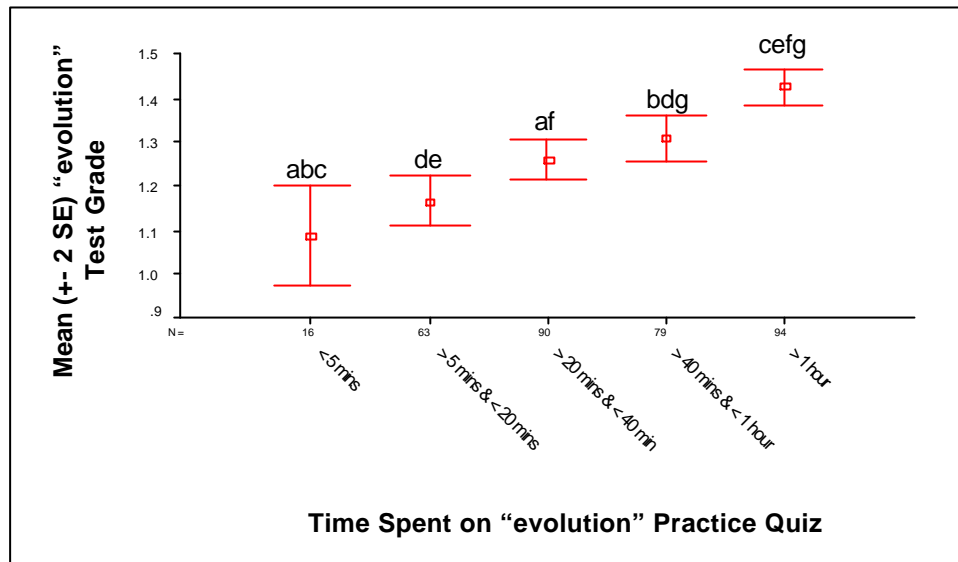
**Table 1. Comments on practice quizzes and graded tests (numbers in parentheses succeeding each comment indicate the number of respondents that made the comment)**

<p><b>Format</b>          Want to know the correct graded-test answers (8)          Some items were confusing (6)          “Choose the correct/incorrect statement” was bad (6)          Sometimes the answers were ambiguous, e.g. labelling on reproductive system diagrams (5)          Would benefit from an explanation of the incorrect answer (4)          Should be able to change graded-test answers before submitting (3)          Separate topics per practice quiz was useful (2)          Keep difficulty of graded tests constant (2)          “Choose correct/incorrect statement” was good (1)          Some grammatical flaws (1)          Graded-test time limit was stressful (1)          Include more diagrams (1)          Confusing to have multiple subtopics per graded test (1)</p>	<p><b>Operation</b>          Some questions were succeeded by/not covered in the relevant lecture/tutorial (15)          Computer troubles, e.g. Internet downtime (5)          The last graded test should be due well before the exams (5)          Graded tests should be available for longer than four weeks (3)          Should be able to do the practice quizzes and graded tests outside the computer room (3)          Group work was unchecked (2)          Closing dates rushed learning progress (2)          Should allow extension times for ill students (1)          Sometimes the closing dates were scheduled too long after the relevant lectures (1)          Noisy computer room was disruptive (1)</p>
<p><b>Learning Value</b>          Motivated revision (16)          Facilitated exam revision (10)          Provided feedback on progress (7)          Graded tests were harder than/different from practice quizzes (6)          Directed study (5)          Practice quizzes were useful for the graded tests (3)          Repeated choices endorsed recall rather than understanding (3)          Generally helpful (3)          Preview of MCQ exam (2)          Multiple-choice graded testing is ineffective (1)          Graded tests were easier than the MCQ final exam (1)</p>	<p><b>Extension</b>          Graded-test closing dates should be announced each week at the tutorial and lectures (4)          Introduce short-answer practice quizzes and graded tests (2)          Introduce a practice quiz and graded test that includes all the topics from the existing six practice quizzes and graded tests (1)          More graded tests wanted (1)          Should suggest related textbook chapters (1)          MCQ exam should be on computer also (1)</p>
<p><b>Philosophy</b>          Graded tests should contribute more to overall grade (5)          Easy marks (1)          Graded tests even out student performance by lifting some weight off the final exams (1)</p>	

***Did use of practice quizzes and graded tests improve general-unit performance?***

Semester-one unit grade increased with an increase in average test grade ( $r = 0.658, p < 0.001, n = 381$ ), number of graded tests attempted ( $r = 0.365, p < 0.001, n = 381$ ), number of practice quizzes attempted ( $r = 0.211, p < 0.001, n = 379$ ) and total time spent on practice quizzes ( $r = 0.297, p < 0.001, n = 362$ ). Independent correlates of semester-one unit grade were average test grade ( $B = 0.464, p < 0.001$ ), number of graded tests attempted ( $B = -0.038, p < 0.001$ ) and total time spent on practice quizzes ( $B = -0.033, p = 0.005$ ), having explained about 47% of the variance. Negative beta coefficients for the number of graded tests attempted and the

total time spent on practice quizzes indicate a negative correlation with the semester-one unit grade. Since all Pearson's correlations were positive, negative beta coefficients were contradictory. However, once the multivariate analysis was re-executed without average test grade, the number of graded tests attempted and total time spent on practice quizzes remained independent correlates of the semester-one unit grade but with positive beta coefficients ( $B = 0.035$ ,  $p < 0.001$  and  $B = 0.012$ ,  $p = 0.001$ , respectively). Hence, whatever factor(s) other than practice quizzes



**Figure 2. Mean (+/- 2 SE) "evolution" test grade by time spent on "evolution" practice quiz (ANOVA,  $p < 0.001$ ). Same letters indicate a significant mean difference. Test grade is arcsine transformed.**

that contributed to the average test grade overrode the relationship between the semester-one unit grade and the use of practice quizzes and graded tests. If general ability was an overriding factor, an effect of the practice quizzes and graded tests on the semester-one unit grade might have manifested for respondents who performed the least well on graded tests. A multivariate analysis was re-executed with the same variables as the original analysis and selection of only the lower third of average test grades. However, total time spent on practice quizzes was not an independent correlate of semester-one unit grade and the number of graded tests attempted had a negative beta coefficient as in the original analysis ( $B = -0.037$ ,  $p = 0.013$ ).

## Discussion

The present evaluation of computer-based practice quizzes and graded tests implemented in a first-year, Human Biology unit at The University of Western Australia provided valuable information about the impact of computer-based assessment on students. Advertisement of the practice quizzes and graded tests was most effective at lectures and tutorials, each a face-to-face medium. Most respondents attempted all practice quizzes and graded tests, consistent with the proportion of students that have used computer-based assessment in other disciplines, such as biology (Peat, 1996) or mathematics (Beevers, McGuire, Stirling & Wild, 1995). Frequent reminders from lecturers and tutors could augment student use of computer-based assessment since some respondents commented that an announcement of each graded-test closing date should be made at lectures and tutorials each week and the most common reason that respondents cited for missing a practice quiz or graded test was "no time".

Almost all respondents endorsed the practice quizzes and graded tests, as was reinforced by the most frequent comment that revision was inspired. Overall, student opinion of the practice quizzes and graded tests corroborated student opinion of computer-based assessment in other disciplines, such as electronic engineering (Sambell, Sambell & Sexton, 1999) or geography (Proctor & Donoghue, 1996). The second most frequent comment was that some questions were followed by or not covered in the relevant lecture or tutorial. However, closing dates for all graded tests followed the relevant lecture and tutorial by at least two weeks and lecturers constructed knowledge statements in reference to unit material. Some respondents made three comments indicating a desire for assessment of higher-order thinking. The comments were that repeated choices endorsed recall rather than understanding, short-answer questions should be introduced and multiple-choice questions were ineffective. Although the practice quizzes and graded tests were designed to assess knowledge rather than higher-order thinking, the three comments prompted staff to consider implementing computer software for assessing short-answer questions, such as E-rater or C-rater (Burstein, Leacock & Schwartz, 2001). However, traditional assessment methods should be implemented where computer software limits staff intentions (Bocij & Greasley, 1999).

Some existing concerns over computer-based assessment (Harvey & Mogeey, 1999) were supported to a small extent. Two respondents commented that group-work was unchecked, indicating that some students may have collaborated on graded tests. The potential for student collaboration is not eliminated with the provision of randomised, open-access, graded tests. However, randomised tests are not identical and so student collaboration may be eliminated if, for example, students sit supervised tests in groups of 20 staggered throughout a week (Thelwall, 1999). Continuous supervision of computer-based graded tests would be practical if students have a designated, computer-workshop session. Five respondents complained of internet downtime, prompting staff to advise students to allow enough time before graded-test closing dates in case computer troubles arise. Four respondents desired more feedback from practice-quiz answers, other than “right” or “wrong”. Staff may want to expand the educational use of the practice quizzes by creating multilevel feedback, a feature of WebMCQ, an Internet-based assessment system (Dalziel & Gazzard, 1999). Multilevel feedback may be in the form of links to information sheets, textbook references or web sites. Only one student requested that the multiple-choice section of the final exam be computer-based also. Hence, staff had little incentive to address the concern over having computer-based continuous assessments but pencil-to-paper final examinations. No respondents expressed difficulty using the computer-based assessment system. Hence, staff may assume that prior computer experience was not an issue for students.

Several respondents made two comments that mirrored the concern of students taking computer-based assessment for Diagnostic Radiology at Sheffield Hallam University (Best, 2002). One comment was that the practice quizzes and graded tests should be accessible outside the computer room and the other comment was that noise in the computer room was disruptive. Since all graded tests were a minor component of the assessment structure for the unit and video monitoring is inadequate for identifying students who collaborate over graded tests in a general access, unsupervised computer room, staff should consider allowing students external access to practice quizzes and graded tests.

Results of statistical analyses were consistent with previous evidence that increased use of computer-based practice quizzes relates to higher performance on computer-based graded tests (Sly & Rennie, 1999; Sly & Stace, 1999). A limitation of the present evaluation was that a measure of general ability external to the unit was not used. Therefore, the possibility that more able students used practice quizzes more was not fully explored. However, a

multivariate analysis revealed that the relationship between use of practice quizzes and graded-test performance existed despite general-unit performance and familiarity with the conditions of graded tests. Further analysis should incorporate tertiary entrance score and/or average grade for other units in the semester. Evidence supports the expectation that general ability will not explain the entire relationship between use of practice quizzes and graded-test performance. Tertiary entrance score did not correlate with use of practice quizzes among first-year, Economics students at Curtin University (Sly & Western, 1998). In a first-year, mineralogy unit at Liverpool University, average performance on other units correlated better with general-unit performance than with performance on computer-based multiple-choice tests (Boyle, 2002). Since the relationship between use of practice quizzes and graded-test performance existed in the present evaluation despite general-unit performance, average performance on other units may not override the relationship. A possible factor affecting the relationship between use of practice quizzes and graded-test performance is whether there was intervening revision. The possible effect of intervening revision can be explored in a further analysis since relevant information was obtained from the questionnaire.

Statistical investigations into whether use of practice quizzes and graded tests improved general-unit performance were inconclusive. Pearson's correlations suggested that respondents who performed better on graded tests, attempted more graded tests and practice quizzes and spent more time on practice quizzes overall, performed better in the unit. However, multivariate analyses revealed that whatever factor(s) affected average test grade reversed the association between general-unit performance and use of practice quizzes and graded tests. Without a measure of general ability external to the unit, a question remains over the effect of practice quizzes and graded tests on general-unit performance.

## Conclusion

The present evaluation of computer-based practice quizzes and graded tests implemented in a first-year, Human Biology unit at The University of Western Australia supported the conclusions that all respondents were aware of the practice quizzes and graded tests, almost all respondents used and recommended the practice quizzes and graded tests, minor improvements to the computer-based system are warranted, increased use of the practice quizzes seemed to have improved graded-test performance and use of the practice quizzes and graded tests was associated with general-unit performance. Further investigation is required to confirm the association between use of practice quizzes and graded-test performance and the nature of the association between general-unit performance and use of practice quizzes and graded tests.

## References

- Beevers, C. E., McGuire, G. R., Stirling, G. & Wild, D. G. (1995). Mathematical ability assessed by computer. *Computers and Education*, 25(3), 123-132.
- Best, H. (2002). *Evaluating the introduction of CBA into the learning, teaching and assessment strategy of the Diagnostic Radiography course at Sheffield Hallam University*. Sixth International Computer Assisted Assessment Conference. Retrieved from [http://www.lboro.ac.uk/service/lt/flicaa/conf2002/pdfs/best\\_h1.pdf](http://www.lboro.ac.uk/service/lt/flicaa/conf2002/pdfs/best_h1.pdf)
- Bocij, P. & Greasley, A. (1999). Can computer-based testing achieve quality and efficiency in assessment? *International Journal of Educational Technology*, 1(1), 1-23.

- Boyle, A. (2002). *A reflective study on the use of CAA to test knowledge and understanding of mineralogy*. Sixth International Computer Assisted Assessment Conference. Retrieved from [http://www.lboro.ac.uk/service/ltd/flicaa/conf2002/pdfs/boyle\\_ap1.pdf](http://www.lboro.ac.uk/service/ltd/flicaa/conf2002/pdfs/boyle_ap1.pdf)
- Bull, J. (1994). *Computer based assessment: Some issues for consideration*. Active Learning. Retrieved from <http://www.ilt.ac.uk/public/cti/ActiveLearning/issue1/bull/index.html>
- Bull, J. & Hesketh, I. (2001). *Computer-assisted assessment centre update*. Fifth International Computer Assisted Assessment Conference. Retrieved from <http://www.lboro.ac.uk/service/ltd/flicaa/conf2001/pdfs/keynote.pdf>
- Burstein, J., Leacock, C. & Swartz, R. (2001). *Automated evaluation of essays and short answers*. Fifth International Computer Assisted Assessment Conference. Retrieved from <http://www.lboro.ac.uk/service/ltd/flicaa/conf2001/pdfs/a.pdf>
- Charman, D. (1999). Issues and impacts of using computer-based assessments (CBAs) for formative assessment. In S. Brown, P. Race & J. Bull (Eds.), *Computer-assisted assessment in higher education* (pp. 85-93). London: Kogan Page Limited.
- Dalziel, J. (2001). *Enhancing web-based learning with computer assisted assessment: Pedagogical and technical considerations*. Fifth International Computer Assisted Assessment Conference. Retrieved from <http://www.lboro.ac.uk/service/ltd/flicaa/conf2001/pdfs/i2.pdf>
- Dalziel, J. & Gazzard, S. (1999). *Next generation computer assisted assessment software: The design and implementation of WEBMCQ*. Retrieved from <http://www.lboro.ac.uk/service/ltd/flicaa/conf99/pdf/dalziel.pdf>
- Harvey, J. & Mogy, N. (1999). Pragmatic issues when integrating technology into the assessment of students. In S. Brown, P. Race & J. Bull (Eds.), *Computer-assisted assessment in higher education* (pp. 7-19). London: Kogan Page Limited.
- Laurillard, D. (1993). *Rethinking university teaching: A framework for the effective use of educational technology*. London: Routledge.
- Peat, M. (1996). *How new technologies are improving our first year biology teaching*. Retrieved from <http://science.uniserve.edu.au/mirror/CUBE96/peat.paper.html>
- Perkin, M. (1999). Validating formative and summative assessment. In S. Brown, P. Race & J. Bull (Eds.), *Computer-assisted assessment in higher education* (pp. 55-62). London: Kogan Page Limited.
- Pritchett, N. & Zakrzewski, S. (1996). Interactive computer assessment of large groups: Student responses. *Innovations in Education and Training International*, 33(3), 242-247.
- Proctor, A. & Donoghue, D. (1996). *Computer based assessment: A case study in geography*. Active Learning. Retrieved from <http://www.ilt.ac.uk/public/cti/ActiveLearning/issue1/proctor/index.html>
- Sambell, K., Sambell, A. & Sexton, G. (1999). Student perceptions of the learning benefits of computer-assisted assessment: A case study in electronic engineering. In S. Brown, P. Race & J. Bull (Eds.), *Computer-assisted assessment in higher education* (pp. 179-191). London: Kogan Page Limited.
- Sly, L. & Rennie, L. J. (1999). Computer managed learning as an aid to formative assessment in higher education. In S. Brown, P. Race & J. Bull (Eds.), *Computer-assisted assessment in higher education* (pp. 113-120). London: Kogan Page Limited.
- Sly, L. & Stace, R. (1999). *Computer managed learning: A new approach with an old tool*. HERDSA Annual International Conference, Melbourne 12-15 July 1999. Retrieved from <http://www.herdsa.org.au/vic/cornerstones/authorframeset.html>
- Sly, L. & Western, D. (1998). Practice tests improve student performance on computer managed learning assessments. In B. Black & N. Stanley (Eds.), *Proceedings of the Seventh Annual Teaching and Learning Forum* (pp. 310-312). Nedlands, Western Australia: Centre for Staff Development.
- Sokal, R. R. & Rohlf, F. J. (1981). *Biometry: The principles and practice of statistics in biological research* (2<sup>nd</sup> ed.). San Francisco: W.H. Freeman and Company.
- Thelwall, M. (2000). Computer-based assessment: A versatile educational tool. *Computers and Education*, 34, 37-49.
- Thelwall, M. (1999). Open-access randomly generated tests: Assessment to drive learning. In S. Brown, P. Race & J. Bull (Eds.), *Computer-assisted assessment in higher education* (pp. 63-70). London: Kogan Page Limited.

## ***Acknowledgements***

The author wishes to acknowledge Jan Meyer for unceasing support and valuable insights spanning the entire project, Julie Hill for regular briefs about the development and operation of the practice quizzes and graded tests, David Coall and Kathy Sanders for useful statistical advice and all participating students.

Copyright © 2003 Frank Gemmiti: The author assigns to HERDSA and educational non-profit institutions a non-exclusive license to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The author also grants a non-exclusive license to HERDSA to publish this document in full on the World Wide Web (prime sites and mirrors) on CD-ROM and in printed form within the HERDSA 2003 conference proceedings. Any other usage is prohibited without the express permission of the author.