

Developments in the provision of quality electronic summative assessments

P. French

The purpose of this Working Paper series is to stimulate discussion and comment. This work may be cited as: Patricia French, "Developments in the provision of quality electronic summative assessments", The Open Polytechnic of New Zealand, Working Paper, March 1998.

Comment on papers is invited and should be addressed to the author/s concerned.

Further copies may be obtained from

Research Manager
The Open Polytechnic of New Zealand
Private Bag 31 914
Lower Hutt

Printed and published by The Open Polytechnic of New Zealand, Lower Hutt.

Copyright © 1998 The Open Polytechnic of New Zealand.

All rights reserved. No part of this work may be reproduced in any form by any means without the written permission of the CEO of The Open Polytechnic.

ISSN — 1174 - 4103

ISBN — 0-909009-07-4

Manuscript submitted for publication in December 1997.

Abstract

This is a Work in Progress paper backgrounding a research area which is to be developed further at the Open Polytechnic. This initial study overviews the written and electronic literature available at 29 June 1998 identifying developments in the field of electronic assessment.

In this study I have focused on ascertaining what was available electronically in summative assessment methods. For some institutions this has simply resulted in the range of traditional forms of assessment being translated onto a computer, while other groups have developed variations available only because of computer usage. This has ranged from the use of electronic mail to transmit assignments between student and tutors to the development of computer adaptive testing. Specifically the study covers summative testing via multichoice questions, constructed answers, essays and practical performance assessments via simulations. There is little information available on summative electronic testing in alternative forms of assessment and so this study does not specifically cover learning contracts, projects, peer assessment, oral examinations, seminar presentations, case studies, laboratory work, or other submitted written work such as literature reviews, book reviews, reports or journalling.

The next stage of the research is to select an appropriate electronic medium for application to an Open Polytechnic programme and to explore the conversion of the assessments in this programme to an electronic base.

Keywords:

Electronic assessment; summative assessment; essay, multi-choice, computer adaptive testing; software; equity issues; simulations

Preface

The issue of electronic assessment has been of interest to me since 1987 when I was appointed to the Nursing Council of New Zealand to be responsible for the Registration Examinations for Nurses. During my time at Council I introduced a total multi-choice examination based on a job analysis of beginning practitioners. This examination utilised electronic capacity in relation to storage and compilation of questions and assessment of the examination. It did not offer electronic administration of the examination. During my time at the Nursing Council, I visited the National Council of State Boards of Nursing in Chicago in 1989 and ascertained the frontier developments they were undertaking in relation to electronic versions of assessment in relation to the State Registration Examination for Registered Nurses. The National Council has continued to be in the forefront of electronic summative assessment as is apparent from the following study.

Contents

Preface	iv
Introduction and parameters of the study	7
Electronic Content Analysis of Essays	9
Computer Adaptive Testing	13
Computerised simulation testing for summative purposes	15
Other Institutional Assessment developments	16
Multichoice usage for assessment of higher level thinking	18
Administering tests on the World Wide Web	19
Authoring software	20
What might all this mean for distance education?	21
Equity	21
Evaluation	22
Issues of Security	23
References	25
Appendix	30
Exclusive Internet usage	30
Formative assessment/online teaching	31

An exploration of developments in the provision of quality electronic summative assessments as at 29 June 1998

Introduction and parameters of the study

The strengths of computer based assessment lie in its potential to provide immediate and accurate feedback to students, the greater potential for flexibility for both students and tutors in types of assessments and the ability of computers to manage all aspects of an examination.

Computerised assessment provides the facility for students to be presented with sound, video animation, graphics, written words and also allows students to input text or manipulate symbols all from the one source and thus does potentially open up broader testing possibilities than a straight pencil and paper assessment. However the need to involve multimedia does depend on the nature of the subject being tested. For example, if an educational objective called for a student to be able to recognise a piece of music this could be incorporated and the student required to identify it and analyse its components. Clearly not all learning objectives require such multimedia use and therefore while the use of multimedia is increasing, most widespread testing is still undertaken via the written language. Such language, itself becomes subservient to technology. This is reflected in the problems which arise when fine nuances and contextually rich scenarios are translated into the logic required by computers. As a consequence it appears many summative assessments of an electronic nature tend to favour multichoice, quizzes, true false, click and drag, short answers or the use of grids. However other more creative developments are occurring.

Computer based assessments are also seen to be of considerable use in the formative assessment area. Here there appears to be great ingenuity displayed in the range of assessments offered. Specifically they have been shown to assist by

- Tailoring learning to the individual
- Providing immediate feedback including complex analysis of student responses
- Allowing the embedding of multiple media within the assessment/ feedback process to create a rich and powerful learning environment.

In this study I have focused on ascertaining what was available electronically in summative assessment methods. For some institutions this has simply resulted in the range of traditional forms of assessment being translated onto a computer, while other groups have developed variations available only because of computer usage. This has ranged from the use of electronic mail to transmit assignments between student and tutors to the development of computer adaptive testing. Specifically the study covers summative testing via multichoice questions, constructed answers, essays and practical performance assessments via simulations. There was very little literature available on summative electronic testing in alternative forms of assessment and so this study does not specifically cover learning contracts, projects, peer assessment, oral examinations, seminar presentations, case studies, laboratory work, or other submitted written work such as literature reviews, book reviews, reports, journalling. However some of the principles which apply to the more standard forms of testing also apply to these latter areas. This study does not look at the area of virtual reality and the possibilities this holds for learning and assessment. Such developments are on the frontiers of computer usage and while needing to be borne in mind are nowhere near entering mainstream computer usage

Electronic Content Analysis of Essays

Some writers consider electronic assessment of constructed answers, particularly essays, only works well if the production of the document is itself part of the test. Other writers consider the technological infrastructure for large scale constructed response assessments is well on its way to being available although technical and practical hurdles still remain. Both groups agree that the assignment must first be produced on computer.

Specific Studies

The assessment of essays by computer was first reported by Ellis Page in 1966. Page's research focused on correlations between simple features of student texts and the grades assigned by teachers. Computer variables were created that approximated values found by human raters. High correlations were found between length of essay and word length while number of common words, spelling errors and average sentence length correlated poorly or negatively. The results showed that the computer program predicted human grades quite reliably or as reliably as human judges correlated with each other. The biggest opposition related to the lack of ability to assess creativity. The important point overlooked by many who opposed this system was that the research was a study of the cognitive processes of experienced essay graders. It could be argued that the research never examined the essays at all but examined what the human graders saw in them.

Several further studies around that time developed the basis of these ideas although there was little interest in using computers in this way. More recently there has been some further work on this area

Idaho University

Hellwig of Idaho State University focused on evaluating business writing. His work built on the Page variables and used a Semantic Differential Scale based on the 'feel of 1000 commonly used words' which were added for a total of potency and evaluation. His research worked well with business reports but an attempt to replicate it on traditional college essays was unsuccessful. Nevertheless his work is interesting in that it raised the possibility of correlating rater judgments with subjective judgments on word choice.

Alaska Assessment Project

The Alaska Assessment Project created an assessment instrument that searched for and counted 24 features. The results were even better than the results achieved by Page. The measures used correlated very highly to the holistic scores (that is the score a human judge assigned an essay). Correlations between the computer variables and human holistic scores ran as high as .96. The variables proved sensitive enough to account for the vast majority of difference between essay grades. However there was a difference in the value given to features such as vocabulary and spelling by different teams of graders. The results of this study show that the software can describe the traits of student writing and the traits of writing evaluators but does not prescribe responses to student essays. The use of such tools as the Alaska Assessment Project instrument can both inform human raters of their tendencies and improve their self awareness and it can provide additional information about student writing but the outcome still required human grading procedures as a benchmark.

Computerised Instrument for writing assessment (CIWE pronounced kiwi)

This instrument was designed in conjunction with the Carmel California Evaluation Centre following work done on the Alaska Writing Programme (it is not clear whether this is the same event as the Alaska Assessment Project mentioned above) To be analysed on CIWE students essays are collected as computer textfiles. This particular study included 85 primary school stories, 82 university year one essays, 243 remedial essays from a community college and 75 mixed university essays. The last 75 essays were compared against teacher scores and a .95 correlation was found. CIWE used factor analysis to show whether any one of the 13 independent variables checked had statistical relationship to any other variables Four factors were identified fluency, sentence development, word use or vocabulary and paragraph development. CIWE processed 40 essays per minute and generated individual and class average raw scores.

Project Essay Grade (PEG)

This is probably the most significant study to date. The Project Essay grade undertaken in 1994 in connection with the Educational Testing Service indicated similar results to the previously mentioned studies. 1314 essays were supplied by ETS that had been composed on computer. 300 of these were

randomly assigned as test essays and 300 were randomly assigned as research essays. No raters were aware of which essay fell into which category. Initially two human raters marked each essay but as this did not give a very consistent result rating for the essays, four further human ratings were done on the 300 test essays and 300 other essays.

Thus 600 essays each received six scores. On the 300 test essays the computer then predicted the average scores of the six human judges. The outcome was that the computer predicted human judgments well possibly better than just the two initial human judges.

While the results of this study showed computer ratings would surpass the accuracy of two judges; would grade more rapidly and economically if the essay was machine readable; and as a result essays could be described statistically, it could not replace human judging. The criterion of success was agreement with multiple human ratings. The results suggested a small sample of each large scale assessment should be rated by humans in order to calibrate the PEG system for a given test. But more importantly the recommendation for further studies in computer essay ratings was to use at least one human rater for each essay.

Objections to essay grading by computer

Objections to computer grading fall into three areas. **Humanist objections** which assert certain choices require human knowledge and background wisdom. However results indicate that the computer results are those most consistent with the judges grades. **Defensive objections** which relate to essay environment such as how to overcome mischievous or hostile students. The PEG programs are so rich in descriptive variables that bizarre elements could be flagged so that any 'odd' essay could be set aside for human inspection. **Construct objections** focus on whether the variables the computer is counting are truly important. This is really the major problem and probably can only be answered by repeatedly undertaking studies such as those above until sufficient evidence is accumulated over a range of topics, even though the basis of reusing human judges is consistency in rating and the computer already meets that criteria.

Revision developments

Developments in revision of work by computers such as spell checkers and proofreading programmes have started to move into the area of checking coherence, development style and tone. Computer programmes can help in revision work. The possibility exists for adaptation of these revision packages into an assessment package.

Assessment of other constructed responses

In 1994 a grid for mathematics tests was introduced in the College Board Science Achievement Tests. This allowed students to generate a response and enter in a grid provided for each answer. Its advantages were the computer could recognise multiple correct responses, the reliability was high because there was no guessing strategy. However it is limited to simple numerical answers without variables.

The Educational Testing Service is also working to adapt symbol manipulation technology for mathematics and science assessment

Computer Adaptive Testing

Computer Adaptive Testing is an example of an assessment available only because of computers, and combines computer technology with Item Response Theory. In these tests the computer performs the tasks of storing, administering, scoring and reporting scores while IRT assesses the item's difficulty, analyzes each student's response, selects the best item to be presented next and compares correctly answered items to their difficulty rating determining the student's ability level. At this present time adaptive testing makes no provision for partial credit. Concerns about the CAT area relate to the establishment of a scale of difficulty for the content domains and item classification which needs strict selection criteria to avoid conceptual overlaps.

There are multiple examples of the application of this form of testing in the literature. The U.S. Educational Testing Service administers it in three high stakes examinations. The Graduate Record Examination the Federal Aviation Administrations pilot certification examinations and the National Council of State Boards of Nursing Registration Examination. There appears to be at least three well known programmes on which to operate this system - CATSYS, BILOG and MicroCAT programmes.

The following is further detail regarding the use of CAT in one the above summative high stakes examination.

The United State National Council Licensure Examination (NCLEX) for Registered Nurses is now offered only as a CAT examination. The priority for the National Council was to ensure the registration exam they offered was psychometrically sound and legally defensible.

The decision to move to CAT was made in 1994 after extensive field tests over two years and a major study undertaken on the legal defensibility of CAT. The field tests revealed that candidates performed comparably on CAT and pencil and paper examinations; that demographic groups seem to be neither advantaged nor disadvantaged; and candidates performance did not appear to be affected by presentation of items on a computer screen and using a keyboard.

Legal issues were addressed in a study conducted by the University of Illinois Centre for Social Research and the legal firm Vedder, Price, Kaufmann and Kammholz. Issues addressed included statistical equivalence, test fairness, review of answers, test anxiety, effects of computer familiarity, test reliability and test dimensionality. The authors of the study's report concluded the court would support CAT.

The validity of CAT for NCLEX was imported from the equivalent pencil and paper test and involved the same item bank. Items in the bank are developed in workshops using detailed item writing guides and review processes, reports of test development panels and committees.

Test plans are based on relationship of the examination to safe and effective practice. This was ascertained by job analyses. The development of a test plan, occurs after each new job analyses.

Testing is available year round 15 hours a day six days a week in five hour time slots. The NCLEX questions are all application based multiple choice involving integrated nursing content which covers four major categories and ten subcategories. There is a maximum time allowance of five hours and a minimum question number including practice questions of 75 with a maximum number of 265. The test terminates when 95% statistical confidence interval is reached about the students competence level. Upon attainment of the 95% confidence level or at the point of maximum time being used or maximum questions being answered the student passes or fails.

NCLEX examinations are available at Sylvan Learning Centres and require students to use only two keys, the space bar and enter/return. All other keys are disconnected. A practice exercise is given prior to beginning the exam

Two types of ID, one with photograph plus an Authorisation to Test (ATT) are required to enter the exam. The name on the ID with a photograph must be the same as that on the ATT. Thumbprint and photograph is taken at the test centre. Photograph is included with the test results. All candidates are observed directly by centre staff and by video and audio recording. The actual data base delivery to testing sites is encrypted.

Computerised simulation testing for summative purposes

Simulations are increasingly available particularly in the science and health sectors but generally the emphasis is on formative testing and online learning.

However, the National Council of State Boards of Nursing are currently piloting clinical simulations for possible use in the NCLEX examination. They consider that CST can capture, time and sequencing of assessment, prioritization of problems, nursing diagnoses, time and sequencing of implementations, reassessment or evaluation following interventions, efficiency of nursing actions (type number and time spent) and risky and inappropriate actions. The CST is based on case studies which unfold over time, pose no questions or answer options and require free text entry of nursing actions. Nursing actions can be awarded either full or partial credit, depending on their correctness, timing sequencing and prioritisation. The pilot study will explore multiple approaches to scoring performance and combining scores across cases. Following the results of the pilot study a decision will be made regarding whether CST should be a component of the NCLEX RN examination from 2001.

It is anticipated the CST software will be available to the nursing community in the near future. Obtaining a sample copy may be very useful in terms of viewing its parameters and potential adaptation to other areas.

Other Institutional Assessment developments

1. TRIADS is a collaborative project between the Universities of Liverpool, Derby and the Open University. The project aimed at ascertaining which learning outcomes can be best assessed using technology and then adapt a generic multimedia, computer based assessment system for formative and summative testing of knowledge understanding and skills. User friendly templates for 18 currently possible question styles would be developed to enable academics to produce their own questions. TRIADs is a three year project and is currently half way through. Some sets of common learning outcomes for areas of geology are complete. The project has extended to nine further higher education institutions. These include Edinburgh, Glasgow, Leeds, Liverpool Plymouth, Sheffield Hertfordshire and Birmingham and areas being developed by these institutions include parasitology, veterinary, biomedical life sciences and learning technology support.
2. The University of Glasgow proposed working on an Interactive Multimedia Computer Based Assessment programme. This was to be funded by the Enterprise in Higher Education Initiative and carried out by University of Glasgow's Teaching within the Independent Learning Technologies programme. TILT. Students were to be tested on wrong ideas included, correct ideas omitted, categorisation, sequencing, describing procedures, deduction and inference. Students would be required to record their answers by multiple selection, true or false, yes or no, single gap fill multiple gap fill, multiple choice, click to select, click and drag to label, rearrangement and selections from labelled grids. The grids could be pictures, words, ideas, equations, formulae, structures, definitions, number, operators or animated movies. Students would chose a cell or combination of cells in response to a question. The number of cells chosen and the order of choice could be made important and used to assess deep learning of high level skills.

The project sounded potentially very useful but upon inquiry the grid aspect of computer based assessment was not undertaken because of other institutions working in computer based assessment although I could find no other institutes considering this particular kind of project.

3. At the University of East Anglia there is project entitled CATS (a computer based assessment of transferable skills project) (currently focuses only on assessment of word processing skills). This project has developed a prototype assessment which allows

Practical examinations of actual performance of a skill
Realistic tasks in terms of content and size
Realistic environment for carrying out tasks
Results analogous with those of human assessors
Information on wide range of performance including content presentation
and analysis of methods
Flexibility in subject matter and assessment criteria

Multichoice usage for assessment of higher level thinking

Killoran: 1992 and Hancock: 1994 indicate that multiple choice can assess Blooms cognitive levels of knowledge, comprehension, application and analysis levels of cognition depending on the learning objectives. However they also considered multichoice was not well suited to synthesis and evaluation or any overt performance oriented objectives such as assessing the ability to write, designing scientific experiments, speaking a foreign language. An information www page originating from Charles Sturt University suggests synthesis and evaluation can be tested via multichoice but gives no research studies to illustrate this.

A study in preclinical courses examined the relationships between course examinations and critical thinking skills, primarily problem solving. The results suggest that objective multiple choice examinations can at least partially reflect critical thinking skills. The study used the Watson Glaser Critical Thinking Appraisal which examined whether the following abilities were checked by the examination. Inference, Recognition of assumptions, deductions, interpretation and evaluation of arguments.

Another article (Karras: 1991), cited Edwin Fenton's research and indicated that Fenton constructed multiple choice tests that explicitly asked student to handle processes of hypothesis building, conceptualisation and other analysis. Karras' article went on to illustrate the application of inferent and deductive concepts in multi choice questions. Such questions test reasoning and a large number of distracters are unnecessary, however the questions required complicated instructions, usually involved false statements and ensuring a 'reason' was completely false at all levels of sophistication was very difficult.

Administering tests on the World Wide Web

Webtester was the only system I discovered that could totally administer tests on the World Wide Web (but there may well be others). All aspects of the examination are computerised and webbased. A web server stores tests, allows properly identified students to take test at prescribed times and places, scores and offers results to instructors and students.

Webtester can deliver tests to anyone in the world via the Internet. Tutors can email tests to this server and set the conditions of the test. Webtester can administer tests, surveys, quizzes and exams. Exams can contain multiple choice, short answer and essay questions. And may include multimedia inserts into any question. As I was not registered as a user of Webtester I was unable to access the exam administration or management section to ascertain the way tests were marked and reported but I inquired directly by email regarding essays and was informed that no electronic assessment of essays occurred - they were sent on to the instructor to do their own scoring. So in effect it merely acted as a repository for the essay questions and a channel for replies for this type of assessment.

The National Council of State Boards of Nursing is in the process of beginning to explore different NCLEX delivery including use of the Internet. However they consider there are some very formidable problems with using the Internet as a delivery device for high stakes licensure examinations. They are interested in maintaining contact on this issue.

Authoring software

The most common software mentioned in the above studies were Question Mark and Authorware. Question mark is a commercial package which includes encryption features to ensure security and limit student access to files. It also has a statistical analysis of test validity feature. Authorware e is a multimedia programme that allows the development of interactive packages. It requires *Shockwave* to be installed to utilise it

In addition the computer assisted teaching and learning Project (CASTLE) aimed to produce a high level authoring shell for online interactive tutorials and assessments in two phases. Phase one resulted in the development of an authoring tool and a mark engine so that tutors could write their own tests for on line use and use them as part of an online student self test or self directed learning package. Since then a 'modify' tool has been written which allows changes to those tests and the addition of multimedia files etc. In addition a 'Create two frame' test is available which enables tutors to combine two web documents to create a two frame test document.

Phase II aims to develop a package so that a tutor can restrict access to a test to a given body of users or for a chosen time period. This would then form part of an online course management package.

Assessor and *Examine* can all deliver tests to higher education students. *Assessor* is a commercial product. *Examine* is an Information Technology Training Initiative ITTI product produced by Nottingham University and is available at minimal cost.

The *Triads project* has produced an engine for processing tests

Tony McCall, Workplace Learning Group, has produced an engine for CAT. It would be useful to involve him in any future developments or assessment of systems for this area.

What might all this mean for distance education?

Electronic assessment appears to be becoming increasingly common. Most of the above studies have involved pilot studies moving to dual mode delivery. Evaluations have been integral to their continued development. Issues also considered by the studies have included equity and security. The current and potential student market in distance education in New Zealand could be expected to respond to the introduction of electronic assessment in much the same way as students in other parts of the world. This means there would be a whole range of responses. Some would consider it of no particular consequence, others would feel threatened. The most obvious area of concern would be that of equity issues.

There already exists equity problems in current forms of assessment. It would appear that computer assessment does not necessarily change any of these and may indeed add some more to consider, particularly, if extensive usage of the computer keyboard or other features is required. Clearly this was overcome in the NCLEX examination by limiting 'live keys' to two. It has been suggested that computers are not neutral pieces of hardware used solely to extend human abilities but should be viewed as embodying particular conceptions of reality, thought and education.

Equity

Appraisals of computer based science assessments suggest the following equity issues. That:

- Using computers automatically constrains the act of assessment in explicit ways. The particular platform chosen will allow only certain interfaces with the test taker, provide limited choices in presentation and enable information for decision making to be packaged in the software in only a few possible ways.
- Using computers will convey benefits to some and impose burdens on others. Anxiety tends to be directly dependent upon the users previous familiarity with the application.
- The design features and modes of computer operation and application, advantage males over females. Studies suggest differences in computer use and attitude may be affected by gender differences within the human brain.

- The wider the range of choices presented to students via computer and other testing formats the greater the problems of reliability, validity and comparability across examinations formats and student selected choices.

Evaluation

An important way of reducing concerns and ensuring credibility would be to introduce electronic assessment in conjunction with evaluation procedures. There are several ways computer based assessment can be evaluated. These include some dimensions that will be obtainable from written studies and others that may need further investigation during or after actual implementation of any pilot study. The following are the range of areas that could be evaluated.

Student, academic and administrative questionnaires,

Structured or semi structured interviews with individuals or groups of students academic or administration staff,

The level of hardware and software needed to support the application system

Comparison of quality of feedback before and after implementation

Comparison of student results before and after implementation

Comparison of number of appeals about results

Comparison of student time spent on assessment tasks before and after implementation

Comparison of costs of development implementation and maintenance of computer versus traditional systems

Accuracy of marking, reduction of clerical errors

Academic and administration staff time saved

Evaluation of visual presentation

Ease of use

Flexibility of application to incorporate change

Level of technical supported needed to maintain the system

Issues of Security

From the administration perspective, electronic assessment would raise issues of security

Firstly there would be the need to identify the person accessing the assessment. A user name password would be insufficient. One option available would be Virtual Private Network with a token system. This requires a physical match between a synchronised number and the system. These are expensive at \$100 a unit/student. Specific machines would need to be set up with this system. Another option is Pretty Good Privacy. This is similar to VPN and could operate with email. It allows verification that the system has not been tampered with through reverse matching arrangements.

Secondly there is security of the assessment site to prevent tampering with the material. While a fire wall keeps people out, it must be broken in order to allow people to access the assessment. Therefore another security measure is required such as Virtual Private Network which allows access through the firewall and encrypts information both ways. This would prevent a person changing information that was there, hijacking a live session or breaking into it or inserting their own commands. Any system should be set up on an institutions own server and so stored on site. This would prevent the storage of the data being accessed at an off campus server site as could potentially occur with WEBTESTER. Electronic assessment would create the need for a bigger communication channel and maybe a bigger server depending on how the server was structured and the size of the database involved.

Thirdly there is the maintenance and further costs of any software improvements. Apart from the purchase and development of appropriate software, there is the staff time involved in producing the products to be put on the software. In particular this applies to the development of an extensive multi choice question bank required for use with any of the MCQ software packages. The model of development utilised by the National Council of State Boards of Nursing offers a high quality, valid, reliable and legally defensible way of producing such an item bank. This involves substantial development time. In addition the introduction of electronic assessment would best be done incrementally as has occurred in most of the studies above were pilot studies and evaluation programmes have been integral to its introduction.

Patricia French
29 June 1998

References

Electronic Essay Assessment

- Hasemer Tone. *Electronic submission of tutor marked assignments*. Electronic Media in Education Research Group 5 June 1996
- McCurry Niki. And Alan McCurry. Writing Assessment for the 21st Century *The Computing Teacher* April 1992 pp.35-37
- Northey Jason. email correspondence regarding student/tutor electronic interface for assignment transmission at Southern Cross University May 1998
- Page Ellis and Nancy Peterson. The Computer moves into Essay grading Updating the Ancient Test *Phi Delta Kappan* 76(7) March 1995 pp.561-565
- Singley Mark and Hessay Taft. Open Ended Approaches to Science Assessment using Computers *Journal of Science Education and Technology* Vol 4 (.1.) 1995 pp.7-20
- Wresch William. The Imminence of Grading Essays by computer -25 years later *Computers and Composition* Vol 10. (2) April 1993 pp.45-58

Computerised Adaptive Testing

- Issues: NCLEX *Using Cat An overview of it all* Vol. 14 (5) Special Edition 1993
- Issues: *Field Tests of NCLEX using Computerised Adaptive Testing: Both NCLEX RN and NCLEX PN Field Tests Completed* Volume 14 (5) Special Edition , 1993
- Issues: Question and Answers: A Guide to the Most Commonly-asked questions of NCLEX CAT Volume 14 (5) Special Edition , 1993
- Issues: *NCLEX Examination continues Psychometric Stability* Volume 17 (4) 1996
- The CAT Testing Environment* <http://www1.kaplan.com/view/article/0,1275,551,00.html>
- Smittle Pat. Computerised Adaptive Testing: Revolutionising Academic Assessment *Community College Journal* August September 1994 238 Interactive media
- Introduction to GRE* <http://www1.kaplan.com/view/article/0,1272,1081,00.html>
- Top 40 GRE Cat Facts* <http://www1.kaplan.com/view/article/0,1275,1084,00.html>

Other Institutional Developments

Boyle Alan. Assessment of Learning Outcomes Outline of TRIADS projects <http://www.pcweb.liv.ac.uk.apboyle/triads/index.html>

Boyle Alan. email message response to questions about TRIADS Liverpool University June 1998

Doughty Gordon and Jane Magill, Ian Turner. University of Glasgow Interactive Multimedia Computer Based Assessment *Active Learning* Issue 1 December 1994

Doughty Gordon. email response clarifying progress on IMCBA project June 1998

Dowsing Roy and Stewart Long, Roman Sleep. University of East Anglia The CATS wordprocessing skills assessor *Active Learning* Issue 4 July 1996

Lloyd Andrew and Carol Holden. University of Brighton Computer aided assessments in pharmaceutical microbiology *Active Learning* Issue 1 December 1994

Higher Order thinking in multichoice Tests

Hancock Gregory. Cognitive Complexity and the Comparability of Multiple-Choice and Constructed-Response Test Formats *Journal of Experimental Education* 62(2) Winter 1994 pp.143-157

Karras Ray. Let's Improve Multiple-Choice Tests *OAH Magazine of History* 6(1) Summer 1991 pp.8-9 and 43.

Killoran James. In Defense of the Multiple Choice Question *Social Education* February 1992 pp.106-108

Leicester University MCQs' and Blooms' Taxonomy <http://www.le.ac.uk/cc/ltg/castle/resources/mcqman/mcqappc.html>

Miller Deborah and John Sadler, Paul Mohl. Critical Thinking in Preclinical Course Examinations *Academic Medicine* Volume 68 (4) April 1993 pp.303-305

Choosing the best type of assessment

Active Learning Number 1 December 1994 Computer Assisted assessment. <http://www.cti.ac.uk/publ.actlea/issue1/index/html>

Bruder Isabelle. Alternative Assessment: Putting Technology to the Test *Electronic Learning* 14(40) Jan 1993

Lord Bill. Choosing the type of assessment and Advantages of computer based assessment Charles Sturt University <http://www.csu.edu.au/division/oli/celt/edtech/assessment/assintro.htm>

Davies David. MedWeb Computer Assisted Assessment <http://medweb.bham.ac.uk/http/cas/>

Coffin D. Using the competitive edge *Journal of Economic Education* 24(1) winter 1993 pp.62-69

Helgeson Stanley and David. Kumar. A Review of Educational Technology in Science Assessment *Journal of Computers in Mathematics and Science Teaching* 12 (3/4) 1993 pp. 227-243

University of Technology Sydney *Assessment* <http://www.clt.uts.edu.au/assess.html>

What kinds of assessment can I conduct on the network? <http://clyde.stac.ac.uk/begininternet/kinds.html>

Internet assessment

Flitton Wayne. email response regarding essay assessment on WEBTESTER 11 June 1998

WebTester Introduction and Description <Http://webtester.weber.edu/docs/into/default.html> <Http://webtester.weber.edu/docs/into/default.html>

Zara Tony., email response to inquiry about CATs, CST and internet assessment National Council of State Boards of Nursing June 1998

Computer simulations

Baker Eva and Harold. O'Neil. Computer Technology futures for the Improvement of Assessment *Journal of Science Education and Technology* Vol 4 (1) 1995 pp.37-45
Software

Coffin Donald. Using The Competitive Edge *Journal. Of Economic Education* Vol 24 (1) Winter 1993

Lord Bill. Charles Sturt University Choosing the type of assessment. Simulations Case Studies Centre for Enhancement of Learning and Teaching <http://www.csu.edu.au/division/oli/celt/edtech/assessmentassintro.htm>

Forney Katherine. Tips for Selecting authoring software Programs *Community College Journal* 64 (2) Oct-Nov 1993 pp.39-42

Issues: Is the Future of NCLEX Examination Computerised Clinical Simulation Testing
Volume 17 (4) 1996

Stephens, Derek. Loughborough University of Technology Using computer assisted
assessment: time saver or sophisticated distraction? *Active Learning* Issue 1
December 1994

Evaluation Issues

Bull Joanna. Computer based assessment some issues for consideration *Active Learning*
Issue 1 December 1994

Open Learning Technology Corporation *Two instruments for Evaluating the Effectiveness
of Electronic Education Systems* [http://www.education.au.edu.au/archives/cons3/
consult3.htm](http://www.education.au.edu.au/archives/cons3/consult3.htm)

Equity issues

Cheek Dennis and Susan.Agruso. Gender and Equity issues in Computer Based Science
Assessment *Journal of Science Education and Technology* Vol 4 (1).1995 pp.75-79

Sutton Rosemary. Equity and High Stakes Testing: Implications for Computerised
Testing *Equity and Excellence in Education* April 1997 pp.5-15

Formative Assessment

Babbitt Beatrice. Hypermedia; making the Mathematics Connection *Intervention in
School and Clinic* Vol 28(5) May 1993 pp.294-301

Lloyd Andrew and Carol Holden. University of Brighton Computer aided assessments
in pharmaceutical microbiology *Active Learning* Issue 1 December 1994

Proctor Amanda. and Daniel Donoghue. University of Durham Computer based
assessment : a case study in geography *Active Learning* Issue 1 December 1994

Weller Herman. Assessing the impact of computer based learning in science *Journal of
Research on Computing in Education* Summer 1996 Vol 28 Number 4 : 461-485

Internet courses - Online Teaching

Resources for On Line Teaching CASTLE <http://www.le.ac.uk/castle/resources/other-sites.html>

CASTLE Project Reports <http://www.le.ac.uk/castle/projectreports/TASC2.html>

Magalhaes Monica and Dietrich Schiel. A method for evaluation of a course delivered via the World Wide Web in Brazil. *The American Journal of Distance Education* Vol 11 (2) 1997

Parrington Norman and Ian. Ferguson. University of Sunderland The use of hypertext and the world wide web in teaching and assessing software engineering. *Active Learning* Issue 1 December 1994

Segal J. An Evaluation of a teaching package constructed using a Web based lecture recorder *Alt-J* Volume 5 Number 3: 32-42

University of the Internet MSC Advanced Microelectronics for Industrialists Press Release 4 March 1988 <http://www.ami.bolton.ac.uk/common/PressBT1.html>

Wood J. Work in progress at the virtual Monash. *Active Learning* Issue 2 July 1995

Appendix

(Information on the following two topics was gleaned incidentally while accessing summative information. It does not reflect any focused attempt to identify the current situation relating to formative assessment/tutoring and electronic media.)

Exclusive Internet usage

The Bolton Institute and the University of Northumbria are running probably the first postgraduate course to be taught entirely over the Internet. All details of the course are accessed (and downloaded) from the Internet by students. Enrolments, teaching, assessment and support are all operated via the internet using ISDN communication links. The ISDN line offers higher speed communication and greater reliability than an ordinary line, it allows voice, data and video to be transmitted simultaneously. This line avoids the need for conversion to analogue form and so provides access to 64 kbps bandwidth

Mgalhaes & Schiel describe the delivery of a pilot course from the University of Sao Paulo. The course was in graphic mechanics which is part of the distance education programme entitled Programa :Educ@r Educ@r project.. The physics programme is delivered via the www and the student registers the results obtained from experiments onto this same web page. This creates an interactive table and the results are verified and returned to the student. At the same time the results are sent to a staff member who is able to comment on them and give additional instructions to students who have had difficulty. The students spends most time on experimental work, calculations and programming and relatively little time sending messages. The first instructor evaluations reveal the following favourable aspects: it provided a stimulus to student understanding of content, allowed actual use of technology, broadened students horizons, facilitated communication and feedback. Improvements were sought with dial up connections and facilities used. Contact with the distance mentor would have been better if it was immediate.

The University of Sunderland offers a course called Software Engineering This course offers lecture and tutorial material in an interactive format and create assessments from a database of MCQs. The entire learning and assessment process is administered via the www. Assessments can be by MCQs marked electronically or by answer forms which are provided on screen and delivered to the lecturer via email. Students can set practices tests for themselves from a second database of practice questions which are automatically marked and results made available to the student. The system uses hypertext. This means

that there is no single order which determines the sequence in which the text is to be read. Hypertext presents several different options, and the individual reader determines which of them to follow at the time of reading the text. As a consequence there are a number of alternatives for readers to explore rather than a single stream of information.

Formative assessment/online teaching

The adaptation of prototypical intelligent tutoring systems available for assessment purposes would be possible. These systems detect deviations in students required performance by consulting a cognitive model of problem solving. As students solve problems with the tutoring interface the simulation of the ideal student model is run simultaneously in the background. At any point in the problem the model should be capable of generating all valid next steps and this should be contingent upon the steps that a have already been performed. If a student chooses a valid step the model remains mute but in the case of a discrepancy the model interrupts with pointed feedback The model can inform the students which of its problem solving rules should have applied and give the result of applying such a rule. Used in assessment contexts the value of such systems may lie in their ability to derive more information from complex performances. I suspect that the CATS project outlined earlier uses an adaptation of this intelligent tutoring system.

Babbit's article indicates that hypermedia can be applied to mathematics to assist in learning problem solving by exploring alternative solution processes. It can also assist in remedial situations to reteach concepts in a manner tailored to meet individual needs.

Coffin indicates successful usage of a simulation program entitled The Competitive Edge aimed at MBA students in the Human Resources area. This program requires students to make decisions about the number of production workers and engineering supervisors to hire, about wages to offer and about product price. The evaluation of this program indicated that it was a valuable and useful simulation of labor market decision making, which added realism to labor economics and labor relations. It was only used as additional course material.

The University of Durham incorporates *Question Mark* software into their formative assessments of geography course for first year students and have found it a useful adjunct in improving factual knowledge.

The University of Brighton utilises computer assessments in pharmaceutical

microbiology which they suggest assists with student centred learning as it offers an opportunity for students to assess revise and reinforce their basic knowledge through competitive interactive assessment. The questions are designed to provide a variety of interactions and involve click areas, multiple selections. Click drag push button and text entry. Uses Professional Authorware on a PC platform.

Monash University uses a computer mediated communication system entitled NETFACE for on line tutor student discussion, student/student discussion and delivery of some classes to distance students. Other NetFace facilities include electronic assignment submission, email, conferencing, library services and Internet access. The NetFace software integrates these functions to present a menu-driven environment

Electronic case studies and problem solving software. A range of this appears to be available focusing on formative assessment or content learning via problem solving. A number of websites show interactive case studies particularly in the health area.

Zlatko Kovacic has done a search of material of cost benefit analysis for online teaching areas as part of his research on this area.

