RE-USABLE ONLINE ASSESSMENT MATERIALS FOR TEACHING ARTIFICIAL INTELLIGENCE

J.E. Smith

Department of Computer Science

University of the West of England, Bristol

James.smith@uwe.ac.uk

http://www.cems.uwe.ac.uk/~jsmith

ABSTRACT

This paper describes the creation and use of a bank of different types of questions for enhancing the teaching and assessment of Artificial Intelligence. The question bank was produced as part of a HEA grant, and has been made publically available as a resource for the community. This paper describes the use of these tools for both formative and summative assessment in a level 1 undergraduate module, and offers guidance based on that experience.

The paper also considers issues relating to the accessibility of electronically delivered assessment to various user groups, focussing in particular on issues arising for students suffering from dyslexia and/or who use screen readers.

Keywords

Question Bank, Assessment, Self-Assessment, on-line materials.

1. INTRODUCTION

This paper is concerned with the creation and use of a bank of different types of questions at various levels for enhancing the teaching and assessment of Artificial Intelligence (AI). This subject has ever-increasing importance in today's "knowledge economy", but is frequently perceived as a somewhat "dry" and "difficult" subject area therefore the development of alternative methods for delivering and assessing this subject is particularly timely.

The creation of these materials was enabled by a grant from the Higher Education Academy under their scheme for "Reusable Learning Objects". Materials.

The primary aims of this project were to:

- facilitate the introduction of AI materials to students.
- assist in the incorporation of AI material into other courses.
- promote widening participation by supporting learning of AI in a structured and self-paced manner beyond the traditional modes of delivery.
- support students through the transition to higher education via the provision of rapid and targeted feedback on their progress and providing guidance on remedial steps.
- address issues of retention and progression via automated means for providing detailed formative feedback available to students, enabling more targeted support based on students' needs, and identification of students experiencing difficulties engaging with the programme of study.

These aims were to be met by the provision of a set of questions that could be used by students and/or tutors to assess understanding of topics in AI. In order to facilitate this process, and in particular the last three points, these questions were created in a format that allowed them to be delivered on-line via a Virtual Learning Environment (VLE). Thus the primary objectives of this project were to create a repository of questions that tutors could rapidly deploy in the form of self-assessment tests covering specific topics, and to monitor their usage. The subsequent decision to evaluate the use of these tools for conducting the summative assessment, was an "added bonus" rather than the primary aim of the project.

The decision to embed these questions within a VLE raised the need to examine issues related to the presentation and accessibility of e-learning materials and how they affect, for example people with dyslexia and partially-sighted users. This is particular important in this field as AI material tends to be jargon and acronym-heavy.

The rest of this paper describes the process and results of this project, and some of the lessons learned.

2. ACCESSIBILITY ISSUES

This work was conducted as part of the HEA project "Creation of OnLine Assessments for Teaching Artificial Intelligence". One of the specific aims of that project was to consider the impact of switching to computer-based delivery of materials for self-assessment on students with disabilities, in particular visual disabilities and dyslexia. A secondary aim was to reflect on existing guidelines on the presentation of materials. To that aim the author worked with the University's Disability Resource Centre (DRC) and the Web-standards group to formulate a set of issues to be considered, and guidelines when creating the assessment questions and test. The short-term and practical nature of the project is such that this report is intended to provide details of the specific issues identified, and of the evaluation of the materials developed, rather than being a more comprehensive review or study of the topic.

It is nevertheless worth considering the topic of whether it is appropriate for all students to use the same sets of learning mechanisms. Clearly in some cases ensuring the accessibility of all materials might diminish the experience for all students, and Kelly et al. have discussed several scenarios in which it may be appropriate to offer replacement learning opportunities [1]. For example, to replace computer-based summative assessment in a clearly time-tabled slot, one option is to provide an oral examination for the people with visual disabilities. However, the purpose of this project was to provide additional tools to complement existing provision, and to enable student-led summative self-assessment. Thus it seems highly appropriate to ensure that any new mechanisms developed do not in fact further widen gaps in learning experience by restricting the ways or times chosen for learning.

It was decided early on in the project to adopt a model where the subsets of the questions developed would be used for weekly formative assessments of a cohort of students, to facilitate rapid adoption of feedback from the users regarding the nature, style, and presentation of the materials. The cohort comprised 103 users, of whom three chose to disclose that they were dyslexic, but none chose to disclose visual disabilities, or that they used screen readers.

In common with many higher education institutes, UWE delivers an increasing proportion of its teaching and assessment materials via a Virtual Learning Environment (VLE) – in this case the proprietary Blackboard system[2]. In order to facilitate rapid adoption by this cohort, it was decided that the materials would be developed and accessed via Blackboard, rather than incurring the time and effort of students familiarizing themselves with another software package.

Meetings with the DRC and Web standards group made it clear that although well known guidelines exist for producing accessible materials [3], many issues remained for users, which can be divided into two groups. The first of these arise from the way that the question setter chooses to prepare and phrase their questions- for example the wording, use of alternative text for images, and the type of questions (multiple choice, true/false, numerical, missing blanks) that are chosen.

The second set of issues arise from the way that the VLE makes those questions available to the users – for example how well screen readers are able to interpret different types of interaction required.

3. METHODOLOGY

Informed by a literature review and initial meetings with the DRC, an initial set of guidelines was produced to be used by the project team (Drs Cayzer and Smith) responsible for drafting questions. These mainly concerned paying particular attention to punctuation and (hidden) html formatting. For this project it was decided to avoid the use of images. The alternative is to make sure that all images are provided with clear and ambiguous alternative text. In some cases this may be unavoidable, but it was not felt to be so for the content being delivered here. Otherwise it was decided to employ a wide range of question types allowed by Blackboard in order to assess whether any of these caused particular problems.

These were used and refined in practice as part of the delivery of a level one undergraduate module "Introduction to Artificial Intelligence". This is a relatively large module (initial cohort of one hundred and three in 2007/8, one hundred and fifty one in 2008/9). The module is mandatory in the first year of the BSc. awards in "Computer Science", "Games Technology" and "Robotics" at UWE, and is an option on the BSc. Computing award. It is also a pre-requisite to further modules in each award such as Symbolic Processing", "Subsymbolic Processing" (both level 2 options, one mandatory Comp. Sci,.), "AI for Games" (level 3, mandatory, Games Tech.), "Machine Learning (level 3 option, Comp. Sci and Robotics). As this is a level one module, it is typically the first exposure that most students have had to the ideas and concepts contained.

Each student is expected to attend a one hour lecture and a ninety minute tutorial per week. The latter has relatively large class sizes, and students work in groups of 5 to 10 creating solutions to problems, presenting their work and documenting it on-line. As well as on-line group documents and discussion boards, the module makes extensive use of various e-learning tools provided Blackboard to deliver additional multimedia content for individual study, as well as detailing the suggested reading from the recommended course textbook [4].

The question bank was used to create formative assessment tests so the students could check their understanding of each week's materials. "Adaptive release" was used encourage participation by making access to subsequent materials dependent on attempting the previous tests.

For ten teaching weeks between September and December 2007 on-line tests were provided and their usage monitored using Blackboard's inbuilt facilities. At the start of the second semester, the staff involved held a meeting to assess the feedback received, and uptake. The findings of this meeting were used to guide the developments of materials for the second semester of the course.

In response to the positive experiences reported by both staff and students involved in this process, it was decided that the summative assessment for the module would also be held via Blackboard. Two sets of forty questions (some of which had been used for the formative assessment tests) were produced – one each for the initial and resit examinations. For practical reasons the students were divided into six laboratories to take the actual examination. Extensive consultations were held with the IT support teams and Academic Registrar to develop procedures that ensured equality of experience and maximized reliability of the system. Detailed written guidelines were produced for the students and invigilators. The exams were delivered as Blackboard "tests", with the settings:

- delivery one question at a time,
- randomize the order in which the questions were presented to each student,
- only allow a single assessment attempt,
- back-tracking allowed,
- no presentation of feedback, or answers submitted after submission.

Once the exam scores had been moderated, a copy of the exam was then made accessible with feedback enabled in order to permit reflection and so that each student could reconsider their responses, and learn from any errors.

4. DESCRIPTION OF MATERIALS

4.1 Technical

The principal outcome of this project is a pool of 107 questions which are available from the url www.cems.uwe.ac.uk/~jsmith/rlo. The question bank is packaged as zip archives of xml files, which can be directly imported into Blackboard, or into other formats by using a suitable xml translator such as those available in Respondus (www.respondus.com). Once imported the question pool may be used to create a range of tests as required. These tests are being continually added to as more materials are developed.

The pool has been designed to provide a wide range of different question types as described in Table 1. The questions cover a range of topics in Artificial Intelligence, and each question is tagged with meta data concerning its primary topic (introduction, philosophy, search, knowledge representation, problem representation, artificial neural networks, evolutionary computation) and level of difficulty (easy, moderate, difficult). They also contain extensive feedback to aid their use in formative assessment. Feedback is provided for correct and incorrect answers, and also where applicable at the level of individual responses.

Most VLEs will offer the course designer to select from a question bank according to different criteria, thus enabling the creation of tests tailored to individual courses. It was not felt appropriate to claim that any particular subset of questions would thoroughly assess any particular learning outcome, as these will differ on a per course/institution basis.

Question Type	Description
multiple choice	User selects exactly one of a series of options.
multiple answer	User selects any number from a series of options. Incorrect selections are penalized.
numerical calculation	User provides numerical response in dialogue box
missing blanks	User provides series of text responses in dialogue boxes to complete sentences.
jumbled sentences	User makes series of selections from pull-down boxes to complete sentences.
matching	User makes series of choices form pull-down boxes to match two sets of items
ordering	User makes series of choices form pull-down boxes to rank a set of items

Table 1: Description of Question Types

4.2 Content

The question bank contains questions to support a first year "Introduction to AI" module, although of course different parts could be used in isolation or at different levels according to the particular award. The questions cover the following topics:

- Philosophy: What is AI?, Turing's test and Searle's Chinese Room argument.
- Learning/Problem Solving as Search: decomposing problems as optimization/model building/simulation using input→model→output version of computing

- Search Strategies: Depth/Breadth first/A*/ Hill climbing. Global/Local search, Search landscapes, Meta-heuristics.
- Knowledge Representation: First order logic, rules, Frames. Semantic Nets, Semantic Web.
- Artificial Neural Networks: perceptron. Usage of simple MLPs.
- Evolutionary Computing: Operation and principles of Holland's Simple Genetic algorithm. Usage of different representations such as permutations, floating point, and tree-format for genetic programming.

4.3 Examples

The following examples show the range of questions types (note they have been reformatted for this paper). One early finding from the weekly tests was that it is desirable to include very explicit instructions with each question, rather than just at the start of a test.

4.3.1 Multiple Choice

Which one of these statements is true?

1. To pass the Turing test a machine must be able to fool a human interrogator each time, for an unlimited period of interrogation.
2. To pass the Turing test a machine must be able to fool a human interrogator most times, for an unlimited period of interrogation.
3. To pass the Turing test a machine must be able to fool a human interrogator each time, for a limited period of interrogation.
4. To pass the Turing test a machine must be able to fool a human interrogator most times, for a limited period of interrogation.

4.3.2 Multiple Answer

Holland's Simple Genetic Algorithm uses Fitness Proportional selection (with replacement) to pick n parents from a population of size n. From these parents n offspring are produced by one-point mutation, and bitwise mutation to become the next generation.

Which of these statements are true about survival of the fittest in this algorithm?

More than one response may be correct, you should identify all that are.

1. The best individual in one generation will always be present in the next.
2. The best individual in one generation will always be present in the set of parents.
3. The best individual in one generation may be present in the set of parents.
4. If the best individual is picked to be a parent, then it will always be present in the set of offspring produced by crossover.

5. If the best individual is picked to be a parent, then it will always be present in the set of offspring produced after mutation.
6. Even if the best individual is not picked to be a parent, it may still occur in the set of offspring.

4.3.3 Matching

This question tests your knowledge of the (Rosenblatt) perceptron and how it learns.

For each of the following scenarios select the correct response from the choices presented and fill in your answer in the table below. You may use the same answer more than once, and some answers may not be needed at all.

Assume this perceptron has 2 inputs x1 and x2, and that the weights w1, w2 and w3 (the bias weight) are all set to 0 (zero).

You are attempting to learn an OR function.

Without ANY training, how would the perceptron respond to each of these inputs and what would be its error? (recall error= TARGET-ACTUAL).

1. x1=0, x2=0   2. x1=0, x2=1

3. x1=1, x2=0   4. x1=1, x2=1

Possible Answers

A. output=0, error=0   B output=0, error=+1

C. output=0, error=-1   D. output=1, error=0

E. output=1, error=+1   F. output=1, error=-1

4.3.4 Multiple Answer

Which of the following statements are true of well-designed Evolutionary Algorithms, and other Computational Intelligence techniques?

More than one response may be correct, you should identify all that are.

It is intentional that the terms "high quality" and "fairly" are not precisely specified.

1. They will reliably produce a high quality solution in a fairly predictable amount of time.

2. They are guaranteed to produce the best possible solution in a predictable amount of time.

3. They are guaranteed to produce a high quality solution in a predictable amount of time.

4. They are not guaranteed to produce a solution of any higher quality than random guessing.

4.3.5 Issues Formulating Questions

One of the early issues that was found when formulating questions was that true/false questions offer a 50% success probability to random guessing, but that it can be difficult to produce as many plausible (yet incorrect) responses as many VLEs require for a multiple choice. In some cases (as per

this example 4.3.1) this is possible. In other we have to avoid using obviously nonsensical answers, and discriminate against pure guesswork. In practice we have tended to join together questions to form a Multiple Answers – such as in example 4.3.4 which tests understanding of two different characteristics of a particular Meta-Heuristic.

In other cases, such as that shown in 4.3.3 there may be a mismatch between the number of responses possible and those needed. While Blackboard and many VLEs allow for some flexibility, as per this example, this calls for precision in the wording of the question.

Finally it should be noted that decision not to use images in questions has caused some limitations. To give an example, if the system presented a partially completed search tree, there would be obvious scope to ask a variety of questions testing the users understanding of the difference between different search algorithms such as breadth-/depth-/best- first, A*, and hill-climbers.

5. ANALYSIS OF BEHAVIOR

5.1 Usage of Weekly Formative Tests

As an ongoing process, feedback was gathered informally in each week's tutorial sessions. It was originally intended to devise and use an on-line questionnaire for the students to evaluate the materials. However in view of the declining usage statistics (see Table 2) it was felt more beneficial to devote a period of the end of term revision tutorials to a discussion of the on-line assessment.

Week	1	2	4	5	6	7	8	9	10
Attempts	91	40	63	31	35	22	17	15	10

Table 2: Analysis of usage statistics for weekly on-line tests in the first semester

To avoid overly influencing the responses, this discussion was allowed to flow, led by the students. A series of questions similar to those in the appendices were held "in reserve" but these were not needed as the topics were covered naturally. Particular attention was paid to the students who disclosed dyslexia, but their responses did not substantially differ from other students. It is perhaps worth mentioning that although most students said they felt the tests were useful means of self-assessment, there was a feeling that because they were not mark-carrying they were assigned a lower priority than written assessments. This may reflect the learning skills of level one students.

This feedback significantly informed the creation of the questions for those parts of the syllabus taught in the second semester. Usage statistics followed a similar pattern – with noticeable drops around the hand-in dates for coursework on other modules.

The usage of the system was monitored on a fortnightly basis, and regular reports made to tutors involved in UWE's Graduate Development Programme (GDP) to assist in pastoral care. This process was reported to be useful by the tutors. The GDP is itself in the process of refinement, and UWE is moving increasingly towards the use of Blackboard. It is hoped that this will facilitate the analysis of student behavior - for example is non-attendance and non-engagement specific to certain modules, or consistent across the first year modules. It is hoped that the ability to automate this analysis will assist in improving retention, increasing participation and monitoring where modules require modification. However, this is still work in progress: Blackboard can produce a wide range of possible reports concerning behaviour on a particular module, and it is not clear which are of most use. For example, during the second semester access restrictions were removed from all of the electronic resources and tests. It then became apparent that many students accessed the lecture notes and tutorials weekly but did not take the tests.

Finally, it appears that many students made significant use of the materials for self-assessment during revision. Of those who attended the examination, over eighty percent had done one or more test within the previous week.

5.2 Additional Testing of Accessibility

To back up this feedback, some additional tests were created to specifically examine the effects of different types of question. These were evaluated in January 2008 by two additional volunteers, both of whom are computer literate, and suffer from dyslexia. Volunteer 1 is a mature first year PhD student. Volunteer 2 is an experienced user of various different screen reader packages, and for this assessment used the software textHELP Read&Write 8.1 GOLD on Windows Vista Business.

5.3 Examination Results

Of the students registered as enrolled on the course, seventy seven sat the initial examination, of which seventy two achieved the required pass mark (40%). One student attended the resit examination but not the original. A further three students attained the pass mark from the resit. Note that this was the only summative assessment for the module.
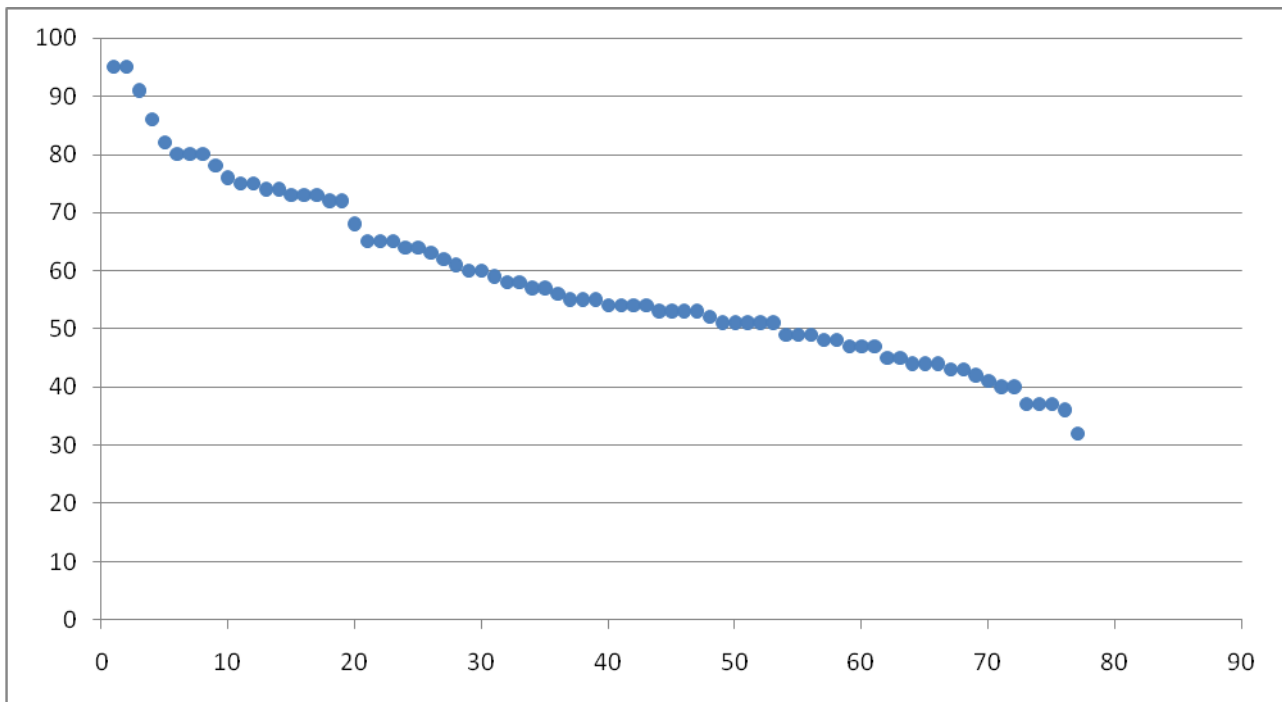
Figure 1: Scatter Plot of Results from On-Line Summer Examination

In order to ensure parity of difficulty when switching from the previous model (2 pieces of written coursework plus a three hour written examination) the electronic exam was moderated by the three members of staff involved in the module, and two other members of academic staff. The marks at the initial exam (5% of cohort failed, 72% passed, 22% "no-shows") represent an improvement in pass rate for the module. In the previous year 79% of the cohort passed the coursework, and 72% the exam, but overall only 48% passed the module at their first attempt, as some students apparently elected to focus on other exams, and then take the resit.

Figure 1 shows a plot of the marks (y-axis) obtained by each student (x-axis) with the students ranked according to mark. As can be seen there is a very even spread of marks suggesting that the on-line assessment provided a good test of the student's understanding of the materials. Analysis shows that the spread of marks was very similar to that seen in the written examination the previous year.

Unstructured discussions were held with the students as they exited the examination, and no students expressed any unhappiness with the way in which the module was assessed. In fact, many expressed the opinion that they preferred the tests to a written examination, as they had been given so much opportunity to familiarise themselves with the style and nature of the questions.

Further analysis is ongoing to see whether the students not attending the examination had in fact withdrawn from their courses.

The system proved fully stable throughout the examination. One student had problems with their machine, but when they accessed a replacement, Blackboard had correctly remembered the exact state of their attempt.

Each question was assigned a number of marks based on its complexity. Table 3 shows the normalised mean score achieved aggregated by question type. As can be seen they are broadly similar except that the jumbled sentence and true/false types appear to be easier, and the "multiple blanks" question scored lowest. Only one of these was used, testing knowledge of specific terms related to RDF and the semantic web. The results bear out the informal feedback throughout the year - that it is almost impossible to second-guess the range of slightly mis-spelt, but otherwise valid answers produced. Note that this module teaches AI, and spelling is not one of the learning outcomes assessed.

Multiple Choice	47.4%
Multiple Answer	49.7%
Jumbled Sentence	74.7%
Matching	53.9%
True/False	73.2%
Multiple Blanks	15%
Ordering	54.1%

Table 3: Analysis of marks obtained by question type

6. ANALYSIS OF ACCESSIBILITY

Throughout the year we used unstructured discussions in tutorials and in Peer Assisted Learning sessions, to elicit the students' feelings about the assessments. We also used additional volunteers to assess the impact of switching to computer mediated assessment for people with dyslexia, and users of screen readers. Both groups expressed the opinion that on-line assessment offered benefits over traditional "paper" methods, particularly for learners with dyslexia. However it has to be emphasised that this benefit only accrues when the setters take particular care in the way that they express and set questions, over and above the established guidelines such as [3,5]. Notwithstanding the small sample sizes, the findings made it clear that certain types of question, such as "ordering" can cause particular difficulties for dyslexic students. More importantly, some types of question may be presented by VLEs in a way that is not properly accessible to users of screen readers. In order to make the learning materials accessible, and in particular if they are to be used for summative assessment, either these should be avoided or specific combinations of screen readers and VLEs should be tested for compatibility before the final assessment is prepared. Further details of these findings, and transcripts of the case studies, may be found in a separate technical report [7].

6.1 Issues of Question Style

In common with many institutions, UWE has guidelines to staff involved in creating on-line teaching materials, but relatively less attention has been paid to how on-line assessment might deal with these issues. Good recent discussions may be found in [5,6]. The former in particular highlights how questions which make use of images or equations need to be particularly careful. Other issues are less obvious, but equally important – for example if different options in a multiple choice questions are not each terminated in a full stop, then many screen readers will simply concatenate them. Many of these issues can be avoided by paying due care to accessibility standards, and so are not treated further here.

One major issue that did arise from the weekly tutorial discussions is that the choice of the type of question used in assessment can have a significant impact. The range of questions types used is described in Table 1. Of these types most students are familiar with *multiple choice*, which is the most common form used as written papers may be easily machine marked. None of the module cohort reported difficulties understanding what was required in the other types of question, although the *matching* and *ordering* question types were predominantly used in the later tests which fewer students attempted, so the sample size is lower.

The feedback from the module students was that the question types did not make significant difference, except for the "*missing blanks*" questions, which were unpopular. This is borne out by the analysis of the exam results (Section5.3) Closer questioning revealed that this was because although a number of alternative spellings could be provided by the question setter, inevitably it is hard to encompass the range of spelling or grammatical mistakes produced. Our recommendation would be that this type of question should only be used where specific technical terms are expected as a response. The decision on whether alternate spellings should be provided and accepted is a matter for institutional policy.

The feedback from the two volunteers was far more instructive here, and may reflect their different experience of education.

- Volunteer 1 commented that "*missing blanks*" questions needed to specify exactly how many blanks were present, and what was expected of the user. On one trial question which started with a "blank", it took them some time to notice this.

- Similarly in "*multiple answer*" questions it was felt that it should be made explicit that more than one option can be selected.

- Both volunteers specifically stated that the "*ordering*" questions were especially difficult for them, which they felt related to their dyslexia.

In general it is clear that when different types of question are employed in the same assessment, then the setter should always include very explicit instructions in each individual question. Since different VLEs may handle and present questions in different ways - e.g. the use of pull-down boxes, radio buttons, tick boxes etc., this means that people wishing to import and use the materials may need to amend the questions to include text relevant to their particular VLE.

6.2 Issues Arising from the VLE itself

The question pools created have been exported from Blackboard as archives in xml format, which can readily be translated and imported into other VLEs. As Blackboard is probably the market leader VLE, this section deals specifically with issues arising from the use of Blackboard 7.1.

6.2.1 Compliance to standards and guidelines.

The Blackboard VLE is stated to comply with the US standard Section 508, and the Web Accessibility Initiative (WAI) issued by the World Wide Web Consortium (W3C). The full statement may be found at [7]. However, it should be pointed out that the web pages produced for conducting assessments do not contain features found in many HE institutional guidelines, such as the options for different contrast settings. To some extent these

could be worked around as questions may be specified in html, rather than blank text, but this would be a time-consuming and inflexible solution. In practice questions are specified without html formatting, and it was left to Blackboard's default settings, which Volunteer 2 commented that they found "Nice and plain, simple font, pretty good". Nevertheless the option to change fonts/contrast would clearly be an advantage for some users.

6.2.2 Implementation of question types.

Both volunteers reported difficulties with the way that Blackboard displays certain question types.

Volunteer 2 reported that the use of pull-down boxes caused problems for screen readers which for "*jumbled sentence*", failed to reliably read the options available. This did not seem to be the case for "*matching*" or "*ordering*" questions ,and appears to arise from the way that the box is displayed.

Both volunteers had problems with the *matching* questions. Volunteer 1 felt the screen layout caused difficulties, and this is reflected in volunteer 2's comments on how the screen reader dealt with it. One option here may be for the setter to explicitly state the options as part of the question text, but this is probably VLE dependent. The questions presented did make use of some complex and fairly long options. The problems experienced by both users suggest that this type of question should only be used when questions and answers can be expressed tersely.

6.2.3 Test Presentation format.

Blackboard offers the choice of displaying the whole test on one screen, or one question at a time. The latter was found preferable for screen readers. It is also preferable for mass assessment, as it reduced network traffic, and each student's test is frequently and automatically saved.

It was reported that the use of capital letters for options rather than numerals was preferred. The exception was for ordering questions, but both volunteers deprecated this type of question anyway.

7. SUGGESTIONS FOR ADOPTEES

There are a variety of ways in which these materials could be incorporated into the delivery of A.I. The tests can be used for both formative (e.g. weekly quizzes) and summative assessment (e.g. end of module exam). The use of meta-data permits the tutor to select questions by type, category and difficulty level as appropriate to their course.

The route followed at UWE was one of quite intensive computerisation of the delivery of content and assessment, coupled with carefully designed large group tutorials. Electronic discussion groups and group work areas were used to provide the possibilities for smaller scale collaborative work and individual communication, and appear to work reasonably well. Students responded well to the clarity and openness of the assessment, and the immediate feedback. On their part, the staff involved reported that it was more enjoyable and productive to spend time throughout the year devising good questions, and providing detailed feedback, than spending the same amount of time marking a large number of similar exam scripts.

Weekly in-class assessment provides benefits in terms of monitoring engagement, but does not fit that well with an agenda of widening access and remote delivery. In contrast, using VLEs to deliver self-assessment materials gains the benefits while providing of materials that students can work through in their own time. Of course students may not be aware when they have misunderstood a concept or topic, and so not see the value of taking a test. Our experience is that to encourage uptake, either the tests should count in some way towards the final mark, or techniques such "Adaptive Release" should be used. In the latter case the materials made available on satisfactory completion of each test must provide sufficient incentive to take the test as many times as necessary.

A vital factor in the provision of e-learning materials is making every attempt to ensure that fair access is available to all students. Some of the findings of this study are noted above. This is an issue that should probably be dealt with on an institutional level, bearing in mind the moral and legal requirements against discrimination.

The following questions are some of the more common of those encountered by the author.

How can this be managed? It all seems like a lot of work…I don't like product X...

On-line assessment can be provided either via customised software, or within the context of a VLE provided at the institutional level. It is this author's opinion that the benefits of using the full range of monitoring and deployment tools available within a VLE – an environment with which most students are having other module content delivered, far outweigh any shortcomings. Developing software, and materials in a customised format may be personally satisfying for the tutor. However it is surely desirable to use software with which the students are familiar, allowing them to focus on the content of the assessment rather than the method of delivery. Moreover, either the accessibility of any custom software has to be properly assessed, or the tutor faces the risk of inadvertently discriminating against some students.

What if I don't have Blackboard?

The question bank developed is supplied in a format which can be immediately imported into Blackboard, and since it is in xml format, can be readily imported into other software via free or low cost tools such as (but not only) the Respondus software.

How can you adequately test understanding and knowledge at level X?

It is possibly the case that simple multiple choice questions may appear to limit the depth of understanding that can be assessed. However, most VLEs now support a wide range of problem types, such as multiple answer, numeric calculation, jumbled sentences, etc, and with due thought these can be used to probe understanding quite deeply. Using these tools undoubtedly requires a rethink of the assessment process, moving to a model whereby the majority of time can be spent refining and developing questions, which are then automatically marked, rather than on marking a number of exam scripts. The fact that this can be done over a longer time period than is traditionally allowed just for the marking process means that any limitations arise more from the imagination of the academic, than from the medium via which assessment is conducted.

8. CONCLUSIONS

One obvious benefit to the academic of employing computer-marked assessment is a reduction in the marking task, enabling fast, or even immediate, feedback. It must be emphasised that these materials provide a way not of reducing the amount of time spent on a module, but of using it more profitably. To give an example, in the next academic year it is intended to extend the range of questions, and provide extra voluntary tests for students who wish to either extend their knowledge of a topic and test it at a deeper level, or to spend more time consolidating their understanding of the topic as taught. Also under consideration are ways of automating the provision of extra materials that students can self-select depending on how they have fared with different topics.

9. ADDITIONAL RESOURCES

A well publicised and significant resource is MITs opencourseware in Artificial Intelligence [9], but this takes a very "traditional" view of AI in its scope, and the assessment tools are not particularly suitable for level 1 students. For more subject specific texts in a range of media, a variety of resources exist, of which most notable is the material on Evolutionary computation in the EvoNet flying Circus[10].

11. REFERENCES

[1] "Implementing a Holistic Approach to E-Learning Accessibility" Brian Kelly,Lawrie Philpis and Caro Howell, UKOKN Web Focus-Alt C 2005. Availabelform www.ukoln.ac.uk

[2] Blackboard version 7.1. http://Blackboard,.com

[3] W3C, Web Content Accessibility Guidelines 1.0 available at http://www.w3.org/TR/WAI-WEBCONTENT/

[4] Coppin. "Artificial Intelligence Illuminated". Jones and Bartlet, 2004.

[5] Online Assessment Tools: Do They Work for Students with Disabilities? Fiedor, L, and Primlani, S. Presented at EDUCAUSE Annual Conferences (10/23/2007). Available at http://delta.ncsu.edu/projects/is/accessibility/assessment/presentation/html/web_data/slides_and_notes.htm.

[6] "Disabled students and assessment through a VLE" B. Newland, J. Pavey and V. Boyd", available from /www.bournemouth.ac.uk/alert/guidelines.htm

[7] Blackboard Accessibility statement: http://www.blackboard.com/company/accessibility.Bb

[8] Smith, J.E. "Analysis of accessibility issues for creating and deploying on-line assessment tools". Technical report available from www.cems.uwe.ac.uk/~jsmith/rlo.

[9] http://ocw.mit.edu/OcwWeb/Electrical-Engineering-and-Computer-Science/6-034Spring-2005/CourseHome/index.htm

[10] http://evonet.lri.fr/CIRCUS2/node.php?lang=fr&