

Analysis of accessibility issues for creating and deploying on-line assessment tools

Dr Jim Smith,
School of Computer Science,
University of the West of England, Bristol

1. Introduction

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 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.

2. Background

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.

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, more than 1 chose to disclose that they were dyslexic, but none chose to disclose visual disabilities, or that they used screen readers. 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 [1], 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.

The rest of this report proceeds as follows:

- Section 3 describes the methodology that was used to conduct this study.
- Sections 4 and 5 consider the two different groups of issues. In both case they begin by identifying some specific issues identified, and how these were intended to be handled. They then summarise the findings of the user evaluations.
- Finally Section 6 provides a summary and overview of the conclusions of this study.
- The two detailed user reports and provided in Appendices.

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 avoiding the use of images, and paying particular attention to punctuation and (hidden) html formatting. 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. For ten teaching weeks between September 2007 and December 2007 on-line tests were provided using some of the developed questions to assess understanding of that week’s materials. Students were encouraged to make use of these tests by having the release of subsequent e-learning materials dependent on taking each week’s test, and usage was monitored using Blackboard’s inbuilt facilities.

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 1 below) it was felt more beneficial to devote a period of the end of term revision tutorials to a discussion of the on-line assessment¹. 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 those of the other students. It is perhaps worth mentioning that all students reported the tests were useful means of self-assessment, but that as they were not mark-carrying they were assigned a lower priority than written assessments. This feedback has significantly informed the creation of the questions for those parts of the syllabus, which will be taught in the second semester.

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

Table 1: Analysis of usage statistics for weekly on-line tests

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.

4. Issues Arising from Question Setters

In common with many institutions, UWE has guidelines to staff involved in creating on-line teaching materials, but there has not been much attention paid to how on-line assessment might deal with these issues. A good recent discussion may be found in [2], which 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 the setters’ 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 2. 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” questions were predominantly used in the later tests which fewer students attempted, so the sample size is lower.

Question Type	Description
multiple choice	User selects exactly one of a series of options.
multiple answer	User selects exactly one of a series of options.
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 2: Description of Question Types

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. 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. The recommendation adopted from this

¹ It is perhaps worth mentioning that all students reported the tests were useful means of self-assessment, but that as they were not mark-carrying they were assigned a lower priority than written assessments.

Analysis of accessibility issues for creating and deploying on-line assessment tools. Jim Smith, UWE.

feedback is that these questions 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 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.

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., to some extent this means that people wishing to import and use the Question pool will need to make sure the questions include text relevant to their particular VLE.

5. Issues Arising from the VLE

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.

5.1 Compliance to Web standards and guidelines.

The Blackboard VLE is stated to comply with the US standard Section 508, and the the Web Accessibility Initiative (WAI) issued by the World Wide Web Consortium (W3C). The full statement may be found at [3].

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 simply would clearly be an advantage for some users.

5.2 Implementation of different question types.

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

Volunteer 2 was 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 be due to the way that the box is displayed (see detailed comments in Appendix 2).

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 explicit state the options as part of the question text, but this is probably VLE dependent. Although 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.

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

Summary

Despite the negative comments above, both the cohorts of students, and the two volunteers 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.

The choice of different question types is a matter of personal taste. Some setters may feel that certain types of question are of more or less use for assessing knowledge and understanding at different levels. Many of the widely known questions types such as multiple choice, and multiple answer were favourably received. Notwithstanding the small sample sizes, the findings above make it clear that certain types of question can cause particular difficulties for dyslexic students. More importantly, it is clear that certain types of question, such as “jumbled sentences” may cause severe problems for 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.

References

- [1] W3C, Web Content Accessibility Guidelines 1.0 available at <http://www.w3.org/TR/WAI-WEBCONTENT/>
- [2] 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.
- [3] Blackboard Accessibility statement: <http://www.blackboard.com/company/accessibility.Bb>

Acknowledgements

This work was funded by the Higher Education Academy development fund as part of the project "Creation of OnLine Assessments for Teaching Artificial Intelligence". The author would like to thank the cohort of students on UWE module “Computational Intelligence” 2007/8, the two volunteers, Martin Serpell, Tom Kendall, Dr Steve Cayzer (HP Labs), and the Disability Resource Centre, UWE.

Appendix 1: Response from Volunteer 1

This user was asked to comment on the different types of question, and the wording.

Here are my comments on the types of controls encountered using Blackboard.

Multiple blanks

It doesn't explicitly say that you need to type something in the box. A blank at the beginning of a sentence is particularly confusing. In my case my brain will try and make sense of a sentence with blanks in it and will try to cope with the missing information, hence it isn't immediately obvious that there is a blank in the sentence. As I said, a blank at the beginning of a sentence caused more of a problem.

Jumbled sentence

Having drop-down boxes made it much more obvious that something needed to be done.

Multiple answers

Must make it very clear that more than one box can be ticked. This was not a problem to me, but this could be that I have designed forms with this kind of input for many years and am therefore familiar with it. I do not know how someone not familiar with it will cope.

Analysis of accessibility issues for creating and deploying on-line assessment tools. Jim Smith, UWE.

Single answers

Using Radio buttons when there is only one answer is great as you get immediate feedback when you try selecting a second option.

Matching

This did cause a problem. Having to look at one part of the screen to see a drop-down box and then look at another part of the screen to see what the drop-down box means is not good. I don't like saying this, but the problem is caused by poor short term memory, which I believe is common in dyslexics. I have to keep moving my eyes between the drop-down boxes and their meanings for every question whereas many non-dyslexics would only need to read the meanings once. This incurs a time penalty and also increases the chance of an error.

Ordering

Need to make it very obvious in which order you want items ordered in. It would be nice if ordering could be displayed more graphically. Dyslexics tend to like pictures more than words.

Lowest							Highest
Dropdown1	Dropdown2	Dropdown3	Dropdown4	Dropdown5	Dropdown6	Dropdown7	

Obviously this layout would not work in all, or even many, cases.

In General

More time should be allowed for dyslexics as they will read slower than non-dyslexics.

Appendix 2: Response from Volunteer 2

For this user some specific question relating to the user of screen readers were produced. These are indicated below. In some places the responses are commented in italics for clarification (e.g. question types).

Q: What software are you using?

I'm using textHELP Read&Write 8.1 GOLD on Windows Vista Business.
I've set it to Web Highlight mode (basically it will start reading from where ever you hover the mouse)

Q: Did the html produced by Blackboard cause problems?

Did not seem to cause any problems (the worst thing in html is badly used tables)

Q: Do you have any preference for all at once or one-at-a-time ?

In one at a time it read "1, 2, 3, ..." After "Question Complete Status:" even when it is hidden (if possible it would be better to have this at the bottom of the page) but this could be avoided by hovering over the question text. At the end of reading the question it read "Moving to another question will save this response. Bar Bar Question 2 of 10 Bar Bar" and then stopped.

In the all at once version it carried on to the next question which might draw your attention to that and stop you thinking about the one you are on.

On balance I prefer the one at a time method

Q: Does the screen reader keep repeating lots of the "background" text or can it correctly focus on just the new bits for each question.

The stuff above and to the left is easy to avoid. The text after it is much harder to avoid but there is not too much. This might be very different for someone using a full screen reader (eg a blind person).

Q: Is it harder to follow a reader when there are lots of options (does the 7 items in memory rule hold)?

I find the reader (along with looking at the screen) so much easier than just reading that it was fine. My problem with the Qs is that I did not know the answers...

Analysis of accessibility issues for creating and deploying on-line assessment tools. Jim Smith, UWE.

Q: Do you feel any particular types of questions would put you at a disadvantage?

I think the Ordering Qs (9 and 10) are probably the worst. I had noticed that I'm always rubbish at fastest finger first on Who Wants to Be a Millionaire. Had not realised it was a dyslexic thing but it makes sense.

Q: Does using random ordering of questions/answers make any difference, or make it harder if you were to redo some questions?

Not sure what you mean.

Q: Does it make a difference if we use letters (A,B,C), numerals (1,2,3), or (I, II, iii) to label choices?

I prefer ABC and find i, ii, iii very unclear. 1,2,3 is probably best for things in order (Q 9,10).

Q: What do you think of the pages visual appearance.

Nice and plain, simple font, pretty good.

Q: Any suggestions for how they could be improved?

Is this part of final marks or just for students to see how they are doing?

In Qs where you type can they ask for a human to look at it and see if it's a spelling mistake?
Points on particular Qs

Question 4 (*fill in the multiple blanks*)

Screen reader stopped when it reached the blank boxes. It could be moved on by hovering over the next bit or by using the forward button (in the screen reader app, see attached)

Question 5 (*jumbled sentence*)

I prefer this set up (no typing allowing miss-spelling) but I could not get the screen reader to read the options. Mostly what it said was Open or Close (as I clicked). Then at one point it did read the selected option but I could not get it to do it again...

The best thing for Q 4 and 5 would be if you could preview the completed sentence and then read it.

Question 7 (*matching*)

On Q7 it always read the selected answer when it got to it and did not stop (unlike Q5). The only difference I could see in the pick lists was that the default was "-" rather than "..."

Question 8 (*matching*)

Just letting the screen reader go from the beginning it read the heading then questions (saying dash) for each pick list and then read the answers. I wonder if it would be better to have the answers above or below the questions. Also rather than just having A-D in the pick lists could you have the full answers?

This was also true for Q7 but I did not really understand it until I'd seen a real example

Q 9-10 (*ordering*)

It allows you to set them all as first. As 9 was in the right order and I don't know the answer to 10 I'm not sure how easy it would be to do for real.

Reviewing the answers

This is pretty good. For Q6 I had selected 3 of the right answers but missed one. It took me a while to see what it was saying I had done wrong as I had 0 out of 10 but all green ticks...

All in all I think it is a pretty good tool and easier for me than a pure paper versions of the same test.