Development and Trial of an Educational Tool to Support the Accessibility Evaluation Process

Christopher Bailey
Teesside University
Accessibility Research Centre
School of Computing
Middlesbrough, TS1 3BA, UK.
+44 (0)1642 384648
c.p.bailey@tees.ac.uk

Dr. Elaine Pearson
Teesside University
Accessibility Research Centre
School of Computing
Middlesbrough, TS1 3BA, UK.
+44 (0)1642 342656
e.pearson@tees.ac.uk

ABSTRACT
This paper describes the design and development of a web accessibility knowledge management tool, known as the Accessibility Evaluation Assistant (AEA) designed to assist novice auditors in the process of an accessibility evaluation. The software incorporates a structured walkthrough approach to guide the auditor through a series of checks for established accessibility principles with the goal of conducting a comprehensive accessibility evaluation. The tool also offers the functionality to tailor the evaluation and prioritise checks based on the needs of two different user groups, or the specific content features of the website. The tool has recently been trialled with a group of 38 undergraduate computing students studying an Accessibility and Adaptive Technologies module with the aim of assessing its reliability and accuracy to validate the method. Some initial conclusions about the reliability and validity of the method and the pedagogical implications of the tool are presented. The results will help highlight the checks that can easily be verified by novices, and those that require a more detailed understanding of accessibility; require informed judgement; or are open to individual interpretation.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Evaluation/methodology.* K.3.2 **[Computers and Education]**: Computer and Information Science Education – *Curriculum.* K.4.2 [**Computers and Society**]: Social Issues – *Assistive Technologies for persons with disabilities*.

General Terms
Measurement, Human Factors, Verification.

Keywords
Web Accessibility Evaluation, Web Accessibility Guidelines, Accessibility Education.

1. INTRODUCTION
One of the contributing factors to the continuing problem of inaccessible websites is a lack of knowledge and understanding on the part of designers and developers and insufficient implementation of techniques to support accessibility [20]. Limited exposure to accessibility during training of I.T. professionals and computing graduates [13] has also been identified as an issue; work is ongoing in institutions to address this by including integrating accessibility topics throughout the undergraduate curriculum [25], however such examples remain exceptions. Unless accessibility is specifically included in the curriculum, computing students will have limited knowledge and understanding of disability and accessibility issues.

As part of their Undergraduate curriculum, computing students within our Institution design and develop rich internet applications and websites. Postgraduate students often carry out authentic projects for 'live' clients. Testing and evaluating their work for accessibility can be a significant issue. The problems the students are likely to face include:

- Limited knowledge and understanding of accessibility guidelines
- Problems interpreting the results of automated evaluation tools
- Very limited access to disabled end users
- No access to expert reviewers
- Little time to dedicate to accessibility in the wider context of their assignments

Projects involving accessibility requirements or working with disabled users have been shown to significantly increase awareness of designing to meet the needs of a diverse population, and students become aware that considering accessibility helps to produce a more usable product for everyone [25]. Undergraduate and postgraduate students increasingly need skills in accessible design to prepare them for work placement projects and employment. They need to understand accessibility beyond the basics, and to apply it in real-world situations, such as developing specific content and solutions for different audiences.

This paper focuses on the development of an educational web accessibility knowledge management tool which aims to address the issue of supporting novice evaluators. It is developed specifically for undergraduate computing students and incorporates a structured walkthrough method to guide the novice auditor through the process of an accessibility evaluation. The development of a prototype Accessibility Evaluation Assistant (AEA) has previously been proposed by the authors [5], since then there has been significant development. It is now a standalone web-based application and has been incorporated into an Accessibility and Adaptive Technology module and tested

W4A2011 – Technical, March 28-29, 2011, Hyderabad, India. Co-Located with the 20th International World Wide Web Conference.

with students. The design of the tool has changed to allow for the inclusion of a separate 'Check Categories' function to enable the auditor to conduct a comprehensive evaluation and support a more holistic approach to accessibility, and video walkthrough tutorials are included (Section 4.0). Evaluation of the AEA was conducted with a group of 38 final year undergraduate students registered on the module and studying a range of computing degrees. The first element of the module assessment (worth 50% of the total mark) was based around the students performing a website accessibility evaluation. The evaluations were designed to test the validity of the structured walkthrough method in terms of its reliability and accuracy, and the students' skills demonstrated through their use of the tool.

The primary function of the tool is to guide the novice auditor through the process of an accessibility evaluation, and assist them in the production of their own evaluation report. Research has identified accessibility evaluation reports as having a positive educational and motivational aspect on those who do not have expertise in web accessibility [20], so a tool which would assist the auditor in conducting an accessibility evaluation was considered to have the potential for strong pedagogical value. An evaluation technique should incorporate a set of easily digestible, specific expert recommendations, tailored to the resource being assessed [17]; this becomes of particular importance when working with novice auditors. The over-riding aim of any accessibility evaluation should be to provide information that is clear, logical and comprehensible to developers [18], so it follows that any tool assisting the process should also be easy to use, self-explanatory and relevant to the context of the site.

The rest of the paper discusses the rationale behind the development of the tool, related work, a description of the tools functionality and structured walkthrough approach, the evaluation methodology and finally, a discussion of the results.

2. RATIONALE

The Web Content Accessibility Guidelines (WCAG) [28], [29] remain the dominant approach in supporting web developers in the development of accessible websites. The limitations of the WCAG approach to accessibility have been well documented [1], [10], [11], [12], [20], [21]. Despite their limitations, guidelines provide an accepted point of reference and large organisations such as IBM produce their own accessibility guideline checklists to test their products [22].

Most of the widely recognised automatic accessibility evaluation tools are based on general purpose guidelines and are not flexible enough to evaluate by specific application type, user group or access device [24]. They can give an indication of the page's conformance level with W3C Guidelines and highlight potential accessibility barriers present but should not be used a sole means of evaluation [11].

Evaluation methods such as conformance review against the WCAG 2.0 Guidelines can be too complex for a novice auditor as they may not have a full understanding of all the issues needed to interpret them, the guidelines need to be applied with informed judgement and the results of any tool used to assist the evaluation need to be interpreted correctly.

Recent work has focussed on the effect of expertise on evaluating for accessibility. Brajnik [8] evaluated the validity and reliability of 21 checkpoints taken from WCAG 1.0 and 2.0 with 35 inexperienced evaluators and found that neither of the guideline sets have checkpoints whose reliability is definitely higher than the W3C recommended threshold. The W3C consider checkpoints to be reliably testable if 80% of knowledgeable evaluators would agree on the conclusion. A study which examined the testability of the 25 highest priority level 'A' success criteria using manual evaluation techniques found that only 8 could be considered reliably human testable when the auditors were novices [4]. These finding are supported by Brajnik [9] who evaluated the testability and validity of WCAG 2.0 with both experts and non-experts. The results for non-experts showed that the agreement level was 6% below that of the experts, they produced 42% false positives and missed 49% of the true accessibility problems. In an evaluation of the Barrier Walkthrough method with experts and non-experts [30], it was concluded that the auditors' level of expertise is an important factor in the quality of an accessibility evaluation. Expert judges were more effective at finding true accessibility barriers and spent significantly less time conducting their evaluation.

One of the deliverables from the Web Accessibility Benchmarking Cluster (WAB) is the development of a standard Unified Web Evaluation Methodology (UWEM) [23] which can be adopted by organisations to assist interpretation of WCAG 1.0 and 2.0. The documentation contains a range of procedures to validate WCAG checkpoints with applicability criteria, expected results for pass or fail and information if the check is fully automatable. The requirement for expertise is recognised in Section 2.4 of the UWEM documentation. We regard the AEA as complementary to the UWEM by supporting novices in gaining a fundamental knowledge of the principles of accessibility by providing a means to support tasks of pedagogical value before using more advanced evaluation methodologies.

Expertise in accessibility is also required to fully appreciate contextual factors when conducting an accessibility evaluation and judging the applicability of guidelines. Context is recognised as having a significant influence in accessibility [7], [19], [12]; in terms of this research, the concept of context is defined as:

- Users of the website with specific access requirements, (e.g. users whose first language is not English)
- Features specific to the individual site,(e.g. Embedded Video clips)
- How the content of the site is presented (e.g. Data presented in tables)

The AEA tool allows the auditor to tailor the evaluation to different contexts, as discussed in Section 4.0.

3. RELATED WORK AND TOOLS

Lopes et al [15] present a Semantic Accessibility Assessment Framework (SAAF) for personalised Web accessibility assessment procedures. The framework has been devised to be applied in different accessibility assessment scenarios so as to separate generic accessibility from domain-specific issues such as evaluating accessibility for different users and device types. Baguma et al. propose a framework for filtering and presenting web accessibility guidelines according to different contexts of use [6]. The MAGENTA tool [14] was developed as a semi-automatic evaluation tool which checks a website against a specified set of

guidelines. The user can carry out an accessibility evaluation from a range of pre-defined guideline sets, and can select which individual guidelines within a set are used. The Accessibility Guidelines Management Framework proposed by Arrue et al [2], is a repository of different sets of guidelines including general web accessibility, as well as those for different application types, end-users and specific user and application type. The framework consists of a web application which allows the user to search for lists of guidelines, create new lists, or edit and update existing lists. OceanAcc integrates an automated evaluation tool with accessibility metrics [16] to provide an application with a semi-automatic evaluation process which simplifies and quickens the evaluation process. User intervention is required to filter false positives and results which are not applicable to the context of the website. A significant development from the ACCESSIBLE EC project is the WCAG 2.0 Web Accessibility Evaluation Tool [26]. This tool allows the auditor to conduct a comprehensive evaluation against WCAG 2.0 checkpoints and tailor the evaluation by impairment or disability type using the ACCESSIBLE harmonised methodology.

The AEA shares some of the same concepts of these tools, particularly the element of context of use, and filtering recommendations by type of disability (or target user group). However, much of the existing work caters for the needs of web developers, IT professionals and other accessibility stakeholders, such as project managers. The AEA is designed as an educational support tool; by guiding the novice auditor through the process of an evaluation, it supports the student in developing skills in accessible web design and development before they use more advanced evaluation tools. The process of conducting an accessibility evaluation can be complex; the proposed tool aims to simplify this process so it can be followed and understood by those without in-depth knowledge of accessibility guidelines.

4. TOOL FUNCTIONALITY AND DESIGN

On launching the AEA [3] the auditor is presented with the option of conducting an evaluation in one of three different ways:

- Check Categories
- User Group
- Site Features

The AEA contains a database of 48 separate accessibility checks based on established accessibility principles taken from a range of accessibility guidelines, established evaluation methodologies proposed by accessibility practitioners and the authors' personal experience of conducting evaluations on a wide range of websites. Many of the checks recommended by the tool require the auditor to manually examine the website or webpage being checked and as such are not suitable for an automated process. The AEA is not an automated evaluation tool, but does utilise existing resources - primarily the Web Accessibility Toolbar 2.0 [27] - to simplify the process of testing and verification.

The Check Categories function, takes the novice auditor through the full set of 48 checks, split into 5 categories. The tool also supports contextual checks for accessibility in two ways:

- By selecting the features of the website they wish to evaluate, the tool prioritises the relevant checks for the auditor, streamlining the evaluation process and eliminating redundancy.
- By evaluating by user group auditors are introduced to the potential accessibility barriers that people with particular disabilities may encounter. The tool demonstrates that by addressing the requirements of a specific disability group, the needs of other disability groups can also be met, but also helps the auditor identify exceptions.

By presenting the relevant guidelines according to a specific context the auditor can carry out an effective audit without the need to go through a full set of checks.

4.1 Check Categories

The Check Categories function has been designed to support a comprehensive accessibility evaluation using all 48 checks – for a full description of the checks please refer to the AEA online [3]. The checks are broken down into five categories making the evaluation process more manageable and grouping related checks in a meaningful way for novices:

- There are eleven Design Checks which examine the visual appearance of a website. Design Checks are concerned with aspects of general presentation, the use of text and colour and the layout and positioning of items. Design Checks are generally conducted by a visual examination of the website, e.g. testing the colour contrast of foreground and background colour combinations used on a page.
- There are fifteen User Checks which are practical, and require human testing. It is necessary to manually interact with the website in order to conduct these checks, e.g. ensuring that navigation elements on a page are accessible using only the keyboard. Although automated tools can help, human intervention is required as some checks are subjective, and require judgement, e.g. ensuring the copy text has an acceptable level of readability and is appropriate for the site's target audience.
- There are eleven Structural Checks concerned with how the content is structured on a webpage. Structural checks also ensure that semantic information about the content of a webpage is provided to the user, no matter how they access a website, e.g. ensuring HTML Heading elements are used to structure the content of a page, and that they are implemented correctly. The five Technical Checks are concerned with coding elements such as successfully validating the HTML and CSS mark-up used to produce a webpage. Technical checks also deal with the metadata elements of a webpage, e.g. specifying a DOCTYPE and HTML language attribute.
- The six Global Checks refer to issues which apply to the entire website, e.g. providing a Site Map and/or Search function or refer to specific functionality, for instance, providing options for user customisation. Those that refer to specific content and functionality can sometimes be verified by examining a single page (usually the home page) and need only be conducted once.

Grouped by category, the checks are presented to the auditor in a list, along with a brief description.

4.2 User Group

The Check by User Group function of the tool currently allows the auditor to prioritise checks based on the needs of 10 different user groups:

- Dyslexia
- Learning Disabilities
- Low Vision
- Screen Reader Users
- Motor Disabilities
- English as a Foreign Language
- Older Web User
- Deaf or Hard of Hearing
- Colour Vision Deficiencies
- Seizure Disorders

The User Group feature is intended to help students relate accessibility to real users and encourage them to find common principles of accessibility which support the development of inclusive interfaces. It is not intended to suggest that separate User Interfaces should be developed for each disability type. With reference to both versions 1.0 and 2.0 of the WCAG, each checkpoint has a fixed priority level based on a level of compliance with the guidelines: A, AA or AAA. It is often the case that issues found when evaluating a site will have a greater impact on one user group when compared to another. For example, if the auditor identified Screen Reader Users as a primary target audience for the site, the AEA would prioritise checks for accessibility issues which primarily affect this particular user group.

4.2.1 Priority Levels

The AEA defines three priority levels for checks:

- Critical Checks are for issues that could render parts of a website or some elements of its content totally inaccessible to the identified user group or may greatly increase the amount of time the user needs to complete a task. An example of a critical issue for screen reader users would be ensuring that content did not rely on colour alone to convey meaning.
- Important Checks are for problem issues which will present accessibility barriers to the specified user group. The user may be able to overcome or find a way around the barrier, but it may cause them a significant inconvenience and will increase the amount of time the user needs to complete a task. An example of a problem issue for screen reader users would be checking that all form elements have a label implemented using HTML.
- Minor Checks are for issues that will present a noticeable accessibility barrier to the specified user group, but the user will be able to overcome, or work around it. An example of a problem issue for screen reader users would be checking that HTML link titles are used only when necessary and do not simply duplicate the name of the link.

The AEA includes a function to compare the checks for two user groups. From an educational perspective, by allowing the novice auditor to filter and prioritise the accessibility checks – and carry out an audit according to the needs of different user groups they will become familiar with the common principles of accessible design, identify specific exceptions and learn about the needs of diverse user groups. From a practical perspective, this tailoring is necessary if the auditor wishes to ensure a website, or a page within a site, is accessible to a specific user group. Students may have to design and develop websites, or sections of sites aimed at specific user groups, but don't have the necessary skills and experience in accessibility or the resources for professional accessibility audits. For example, individuals with learning difficulties represent a user group where standard accessibility guidelines alone may not be appropriate as they do not cater for symbols based interfaces and resources, which may be the most appropriate way to present content to this user group.

4.3 Site Features

This function allows the auditor to filter the checks based on specific elements of a website, namely:

- Forms
- Images
- Cascading Style Sheets
- Links
- Multimedia
- Semantic HTML
- Tables

When using a conformance review evaluation methodology which organises checkpoints into a number of priority levels, such as WCAG 2.0, accessibility guidelines for a single element or site feature (e.g. forms) can be spread across two or three different priority levels. This could be confusing for a novice evaluator and makes the evaluation process overly complex. This complexity can be addressed by grouping checks together based on the element or site feature they refer to, thereby increasing the relevance of the checking process and eliminating redundancy.

4.4 The Structured Walkthrough Approach

Having selected checks based on Categories, User Group or Site Feature the second element of the AEA provides a step-by-step walkthrough for each check. The structured walkthrough approach to evaluation is based on the Barrier Walkthrough method [7]. For each accessibility check the auditor is given support on:

- The accessibility principle they are checking for
- The user group(s) affected
- The nature of the barrier or problem caused
- A procedure for checking and verifying the issue

The procedure for checking and verifying may be manual, automatic or a combination of both. Where the check directs the user to an automated check or functionality provided by the Web

Accessibility Toolbar, in addition to being instructed which element or menu to use, advice is given on interpreting the results; this is considered to be one of the key elements of the AEA as an expert system.

A short video demonstration of an expert evaluator performing the check is also provided. This includes a commentary describing the check procedure, highlighting the accessibility barriers found, and gives advice on interpreting the results of the automated elements of the Web Accessibility Toolbar. An example of the typical instructions provided for the auditor – in this case for checking image text alternatives – is provided in Figure1 below.

For each check the user is given four options and can select whether the check is; Met, Not Met, Partly Met or Not Applicable. There is also space for the auditor to enter comments, and once they have completed the evaluation, they can print a copy of the report.

Image Text Alternatives
Check that all images, and similar elements, have an appropriate text alternative that accurately and concisely describes its content and/or function.

Why this is important
Text alternatives are important for screen reader users as the text is read aloud by the software. If written properly they describe the content or function of an image. They also act as a tooltip as some browsers display the text alternative when the user hovers over the image. A null text alternative of empty quotation marks can be used if the image is purely decorative as this will instruct the screen reader to ignore the image.

How to check this
The Web Accessibility Toolbar can assist with this check but it must be manually verified:

1. Select "Images" > "Remove Images"
2. Images will be removed from the page, and the text alternative will be displayed
3. Where there is no text alternative a warning of "No Alt!" will be displayed

To verify the check:
1. Check and record if all images have a text alternative
2. Check that the text alternative is concise, accurately describes the content of the image and is related to the content of the page
3. If the image acts as a link (or has a function) the text alternative should state the function or page it links to
4. If the image is purely decorative, is used to add visual appeal to the page or is a spacer image, check that it has a null text alternative

Figure 1: Example Instructional Information

5. TRIAL

The AEA was trialled with a group of 38 final year undergraduate students studying a range of computing degrees. The students had limited previous knowledge of web accessibility and none had conducted an accessibility evaluation. Contact time with the students was two hours a week; a one hour lecture, followed by a one hour tutorial, with additional support available upon request. Over the course of the module the students were introduced to a range of topics related to accessibility, including types of disability, disability legislation and legal case studies, WCAG Guidelines, assistive technology and evaluation methods and tools. The students were introduced to the AEA application in the fifth week of the module, one week before they received their assignment during which they would use the tool.

There were two separate elements to the students' assignment which were designed specifically to test the validity of the structured walkthrough evaluation method, and the effectiveness of the AEA as a learning support tool. The students were instructed to carry out an accessibility evaluation of a given website using the Check Categories function of the AEA; assess the site features which support accessibility; and identify any potential accessibility barriers. Four websites were used for this study, and each student was assigned one from:

- Vancouver 2010 Winter Olympics website: http://www.vancouver2010.com
- CNN: http://www.cnn.com/
- The F.A. Premier League: http://www.premierleague.com/
- WalMart: http://www.walmart.com/

The websites were chosen for their size, complexity and diverse range of content. An evaluation of the Home Page was compulsory. The students made their own choice of the other 2 pages but were instructed that the three pages together must contain the following content features:

- Multimedia or Flash content
- A page containing multiple (10 or more) Photographs, Images or Graphics
- Form elements
- Navigation Menu items
- Table elements

This was to ensure that the students used the full range of checks contained in the AEA. For each check they were instructed to make a judgement from the four options (Met, Not Met, Partly Met or Not Applicable) and were required to provide a short justification of their decision. This would assist in the analysis of the result by helping to identify false positives, erroneous decisions or cases where the student had misunderstood the requirements for the check.

The second part of the assignment was designed to assess the students understanding of accessibility in relation to specific user groups through the use of the AEA. Each student was given two (of a possible six) personas describing the needs and preferences of hypothetical users based on the groups contained in the AEA. Personas give a detailed, realistic description of a typical user. The use of multiple personas is intended to help students to relate accessibility to real people and encourage them to find common principles which support accessibility for a range of users. Each persona was not necessarily related to a single user group. An example persona is provided in Figure 2.

Persona 2: John, 70

John is 70 years old and lives with his wife in Yarm. He worked as a structural engineer for many years but he had to take early retirement due to ill health. When he was 50 he started to develop cataracts and his vision has slowly deteriorated since then, although he can still carry out most everyday tasks. He and his wife both have arthritis so they have begun to spend more time at home. John often used a computer when he was working but has never had one at home. John's son, David, is about to move to America with his wife and daughter, so John has just bought a new PC with Windows 7 so they can more easily keep in touch. He bought it so they could keep in touch more easily via e-mail and Skype however David has been showing John how he can do lots of things online such as banking, shopping and keeping up-to-date with all the latest news and sports results. John was having some difficultly seeing everything clearly on the monitor so David installed ZoomText (a screen magnifier) on Johns' PC. Unfortunately, David installed an old version of ZoomText which isn't 100% compatible with Windows 7 and although John can use it, it sometimes crashes. John can use a mouse, but the symptoms of his arthritis also mean he can't use the mouse for long periods. David has started to look for something more appropriate for his father.

Figure 2: Example Persona

The other five personas were presented in a similar manner to Figure 2 and represented a screen reader user, a person with dyslexia, a person with motor disabilities, a user with low-vision and whose first language was not English and an individual with learning difficulties. The students were instructed to use the AEA User Group function and write a short reflective piece identifying how the website meets the needs and requirements of the two different users (or personas), and identify potential barriers they may face. They were encouraged to consider the Critical and Important checks identified by the AEA for their personas and discuss the similarities and differences between the checks.

By comparing the results of the students' evaluations, the reliability of the AEA could be determined. Reliability refers to the extent to which different evaluations of the same page lead to the same results. An expert auditor also conducted an evaluation of the home page of each site using the same structured walkthrough method in the AEA. By comparing the results of the expert evaluators with that of the novices, conclusions could be drawn about the validity of the AEA. Validity refers to how well novices can accurately identify the accessibility barriers present on a page [8].

In order to get feedback on the effectiveness of the AEA itself as a tool to support understanding of the evaluation process, the students were asked to reflect on their own personal experience of carrying out the accessibility evaluation, describe any problems they encountered and indicate how the process might be improved.

6. RESULTS

The preliminary analysis of the results of the students audits focus on two aspects of the process. The first is a comparison of the extent of agreement between the novice auditors for each check to examine the level of consensus and evaluate the reliability of the method. The second is a comparison of the decisions made by the novice auditors with the expert evaluator. For the purpose of this paper, only the home page is considered as this was common to all student evaluations.

6.1 Novice Consensus Summary

Vancouver Olympics Home Page: 96% (46/48) of the checks had a consensus agreement of 50% or greater, 79% (38/48) had 60% or greater and 60% (29/48) had 70% or greater.

CNN Home Page: 83% (40/48) of the checks had a consensus agreement of 50% or greater, 66% (32/48) had 60% or greater and 50% (24/48) had 70% or greater.

Premier League Home Page: 71% (34/48) of the checks had an agreement of 50% or higher, 58% (28/48) had 60% or higher and 45% (22/48) had 70% or higher.

WalMart Home Page: 79% (38/48) of the checks had a consensus agreement of 50% or greater, 67% (32/48) had 60% or greater and 56% (27/48) had 70% or greater.

When combining the overall results of the four Home Pages 82% of the total checks have a consensus agreement of 50% or more, 68% have a consensus agreement of 60% or more and 53% have a consensus agreement 70% or more. While these figures do not meet the reliability level of 80% required by the W3C for knowledgeable evaluators, given that the auditors were accessibility novices and completing their first evaluation, we consider these figures promising.

6.2 Expert Comparison

The result of the novices' evaluation was compared with that of the expert auditor.

***Vancouver Olympics Home Page*:** 75% of the checks (36/48) had a decision with which 50% or more of the novices matched that of the expert auditor. Overall, 65% of the total decisions for this page matched that of the auditor.

CNN Home Page: 71% of the checks (31/48) had a decision with which 50% or more of the novices matched that of the expert auditor. Overall, 60% of the total decisions for this page matched that of the auditor.

Premier League Home Page: 52% of the checks (25/48) had a decision with which 50% or more of the novices matched that of the expert auditor. Overall, 56% of the total decisions for this page matched that of the auditor.

WalMart Home Page: 60% of the checks (29/48) had a decision with which 50% or more of the novices matched that of the expert auditor. Overall, 58% of the total decisions for this page matched that of the auditor.

Taking the results of the four home pages together, 65%, or 124 of the 192 checks in total had a decision with which a minimum of 50% of the novices matched the decision of the expert. Overall, 60% of the total decisions for all pages matched that of the auditor.

An example of the full results of the evaluation for one of the Home Pages (Vancouver 2010 Olympics) is found in Table 1. The

table shows the percentage of decisions made for each judgement; where the novices' decision matches that of the expert, the percentage is marked in bold and shaded yellow.

Table 1: Results of Vancouver 2010 Home Page Evaluation

Design	Met	Not	Partly	N/A
1	-	**70%**	30%	-
2	**40%**	-	60%	-
3	**30%**	20%	50%	-
4	**70%**	-	30%	-
5	10%	**60%**	30%	-
6	10%	10%	10%	**70%**
7	30%	**50%**	20%	-
8	10%	20%	**70%**	-
9	**70%**	-	30%	-
10	**90%**	-	10%	-
11	**20%**	20%	30%	30%
User	**Met**	**Not**	**Partly**	**N/A**
1	-	**100%**	-	-
2	70%	-	**30%**	-
3	20%	**30%**	50%	-
4	**10%**	60%	20%	10%
5	-	30%	10%	**60%**
6	10%	70%	**10%**	10%
7	**60%**	-	40%	-
8	20%	**50%**	30%	-
9	-	20%	-	**80%**
10	10%	**20%**	-	70%
11	-	**90%**	-	10%
12	-	**90%**	-	10%
13	-	**70%**	30%	-
14	40%	50%	**10%**	-
15	-	**20%**	20%	60%
Structural	**Met**	**Not**	**Partly**	**N/A**
1	**100%**	-	-	-
2	**80%**	-	10%	10%
3	30%	10%	**30%**	30%
4	10%	50%	10%	**30%**
5	-	10%	-	**90%**
6	**90%**	-	10%	-
7	-	30%	-	**70%**
8	10%	**70%**	10%	10%
9	-	**70%**	10%	20%
10	10%	**60%**	20%	10%
11	20%	20%	**60%**	10%
Technical	**Met**	**Not**	**Partly**	**N/A**
1	**80%**	-	20%	-
2	10%	**60%**	20%	10%
3	**90%**	10%	-	-
4	-	**80%**	20%	-
5	-	**70%**	30%	-
Global	**Met**	**Not**	**Partly**	**N/A**
1	-	**90%**	-	10%
2	10%	**90%**	-	-
3	-	**100%**	-	-
4	10%	10%	**80%**	-
5	40%	10%	**50%**	-
6	**100%**	-	-	-

7. DISCUSSION OF RESULTS

By examining the students' justification of their decision, we are able to draw some conclusions as to why there were inconsistencies in the judgement decision. Some of these explanations are similar to the study conducted by Alonso et al [4] who found that knowledge, effort and comprehension affected their decisions when evaluating against WCAG 2.0 Guidelines. For many individual checks there are more than one criterion, for example, a site map should be provided and the link to it placed in a prominent position on the page. If the website meets one of these requirements, but not the other, then the auditor could reasonably justify a judgement of met, partly met or not met. In our experience of producing professional reports we tend to highlight specific examples of instances where accessibility barriers are present, and guidelines have not been followed. For example, isolated instances of form labels not implemented using HTML elements, would produce a result of 'partly met' rather than stating that a website 'fails' a specific aspect of accessible design.

The requirement for justification does highlight the effectiveness of the method; it encourages the students to consider the subtleties of the check, take into account the results of other checks and consider the check in the wider context of the page. This was particularly evident in checks for Image Text Alternatives, the judgement was classed as met, partly met or not met depending on the number of images missing an alternative and whether the auditor considered these important or not.

There were also examples of the evaluator effect, as discussed by Brajnik [8]. Based on experience, the expert auditor might ignore what may be considered very minor accessibility problems. For example, if a small number of unimportant images were missing a text alternative or a commonly used acronym (e.g. U.S.A.) that was not expanded using HTML elements the expert would class the check as Met, whereas the novices judged the elements more strictly and classed it as Not Met or Partly Met. This was also evident when examining a rotating Flash Banner on the Vancouver website. This caused many students to judge the Moving Elements check as Not Met, as technically there was a moving element on the page, however the expert judged it as met, as the user did have control over which part of the banner was displayed.

Many of the novices misinterpreted checks, for example, they discussed colour contrast in Use of Colour check (which refers to using colour as a sole means of identifying information) when a colour contrast check features separately in the evaluation process. The check for Link Titles was often misinterpreted, for example, the novice classed the check as Met if link titles were present, when the check actually refers to unnecessary duplication. This may be a result of the student misunderstanding the instructions in the AEA tool.

Students often also often used Not Applicable and Met to signify that there was no accessibility barrier, for example, if there were no moving elements on a page, the student would make a decision of Not Applicable when the expert classed as Met. Both judgements and justifications were correct, and as long as there was no accessibility barrier this is acceptable, although it has skewed the initial results.

The differences in decision was also sometimes due to lack of knowledge or carelessness – the student had not examined the page correctly and had missed the issue, for example, many students missed the lack of text alternatives for buttons designed as graphics, or had missed a site map when one was present, but not prominently displayed.

Students sometimes took examples provided in the AEA too literally, for example, the Supplementary Images check suggested that a flag could be used to signify that the site was available in different languages, and said the check was Not Met if this was not present. Another example was the Required Form Fields check; the novices judged it as Not Met when a login feature with form fields for a Username and Password were present on a page. Sometimes the justification did not match the judgement decision, for example, judging a Valid HTML check as Met even though there were a large number of validation errors, perhaps misinterpreted as *is it possible* to validate the page. If the result was recorded incorrectly it may indicate a limitation of the tool as it was not clear to the student what exactly was required of them or what they were supposed to record. This was sometimes supported by evidence in their personal reflection of the evaluation process but in some cases it was not possible to understand why a student had reached a decision as they did not supply sufficient justification. In some cases the novice was simply incorrect, in one example they had classed a check referring to the accessibility of Flash elements as Met, as Flash content was present on the page.

As with any study involving students, the results are affected by the amount of effort and time each individual put into their individual assignments. Although some students had hurriedly completed the assignment, even giving the date of evaluation as the day before the assignment deadline, many conducted a thorough evaluation and in a few cases found issues which the expert missed, for example a thumbnail gallery with a sequence of images.

Many checks are simply open to interpretation where the checks are subjective, e.g. judging when instructional information is required or whether navigation links are big enough. This could be considered a limitation of the tool if clearer guidance could be provided.

7.1 Student Feedback

There was a general consensus among the students that the evaluation process took a significant amount of time. However it is important to note that the tool does break down evaluation into a series of checks and check categories which do not have to be completed at once. As it was the students' first evaluation, it is reasonable to expect that it would take a considerably longer than the expert auditor. Once the students become familiar with the AEA, the evaluation process, and become more experienced evaluators, it is expected that the amount of time required would decrease.

Although the tool is organised by types of check, some students' indicated they would prefer it if similar checks were grouped together. Although such a grouping exists in the Site Feature function of the tool, this was not evaluated by the students. It may be beneficial to repeat sub-groups of categories within the main Check Categories. The students commented that the evaluation process was well explained, the progression through the checks was clear and the AEA used language they could understand. Many commented that they found the video tutorials useful as it showed them examples on live websites. Students indicated that they found making judgements about some checks such as colour contrast and text alternatives difficult and suggested they needed to conduct more evaluations to gain experience.

One possible limitation of the tool is that the classifications are not clearly explained – however it is difficult to provide a blanket definition for each of the decisions as a judgement must be based on the specific context in which an accessibility barrier is found, and the results of other checks. Some students commented that at times it was difficult to judge the effect a check would have on someone with a disability. They believed with more research into each disability and the needs of each user group the evaluations would be easier to complete and the ability to distinguish between the classifications would be easier to decide.

There was evidence that the tool increased the students' understanding of the needs of different user groups. Our initial analysis suggest the user group function of the tool supports the auditors in making decisions about the relative impact of accessibility barriers on different user groups. Students used the personas to support their evaluation and assessed accessibility barriers in relation to their personas. The full analysis of this aspect of the tool will be the subject of another paper which will focus specifically on the use of profiles to provide context for novice evaluators in what might otherwise appear a purely abstract exercise.

Some of the students' comments illustrate their experience of the tool:

"From evaluating using the personas it really shows that you have to consider that websites will have a lot of user groups that will visit the website and will expect certain requirements to be met. I thought that the AEA was a great application to use for this report as it guided me through the evaluation process – I will certainly use this method again."

"This was the first time I have completed an accessibility evaluation and I found it opened my eyes to a lot of issues that I wouldn't have considered before. I found the whole process simple to follow using the AEA tool and it made it very simple when unsure about how to evaluate a certain point, I now understand the problems that would be faced by someone with a disability and I understand how I could design a website to limit these problems."

"I did not know there were so many considerations when developing a website – the AEA made realise that there is more than just producing valid code."

"Carrying out the accessibility evaluation really helped me understand the importance of web accessibility. Before I carried out the evaluation, and when designing websites myself I didn't take most of the checks into consideration – mainly because I didn't know what checks to carry out."

8. CONCLUSION

While it is difficult to draw firm conclusions from a single study, the initial results indicate both the potential of the structured walkthrough method as an evaluation tool and the AEA as a

learning application to support novice evaluators. Although initial results are presented in this paper, a deeper analysis will explore the extent to which students failed to detect true accessibility barriers and were wholly incorrect in their judgement. The tool does demonstrate potential in an educational context. Requiring students to articulate and justify why they have made their decision encourages them to consider the accessibility implications of each check in more detail and assists the tutor in giving feedback about potential erroneous decisions.

It is also worth bearing in mind that this was the first evaluation that the students conducted; it would be useful to examine the results if the students conducted several more evaluations to see if the reliability and overall effectiveness of the evaluations increased. In terms of evaluating the validity of the method, it would be interesting to compare the results with evaluations conducted by expert evaluators. Further analysis is also required to compare the results with studies by Brajnik [8] and Alonso [4] who evaluated the 25 Level 'A' success criteria of WCAG 2.0 with novices.

One aspect which was not considered as part of this study (as it was conducted as part of the students' module assessment) was the actual time taken to conduct the evaluation. For future studies it would be interesting to explore the time taken by each individual auditor to conduct an evaluation, compare this across different websites and measure any reduction in duration in consecutive evaluations.

Overall, the study adds further evidence that accessibility evaluation requires expert judgement, although the results of this first trial indicate that some of this expertise can be incorporated into the AEA. We are encouraged that by incorporating the tool into the teaching of accessibility, students can be supported in developing skills in accessibility evaluation.

9. REFERENCES

[1] Abascal, J., Arrue, M., Fajardo, I., Garay, N., and Tomas, J. 2004. The use of guidelines to automatically verify Web accessibility. Univers. Access Inf. Soc. 3, 1 (Mar. 2004), 71-79.

[2] Arrue, M., Vigo, M., Aizpurua, A. and Abascal, J. 2007. Accessibility Guidelines Management Framework. *In C. Stephanidis (Ed.). Universal Access in HCI, Part III, HCI International 2007* (Beijing, China, July 22-27, 2007). LNCS 4556, 3--10, Springer, 2007.

[3] Accessibility Evaluation Assistant. *Accessibility Research Centre.* http://arc.tees.ac.uk/aea/. Accessed: 27/11/10.

[4] Alonso, F., Fuertes, J.L., Gonzalez, L.A. and Martinez, L. 2010. On the testability of WCAG 2.0 for beginners. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)* (W4A '10). ACM, New York, NY, USA. DOI=10.1145/1805986.1806000 http://doi.acm.org/10.1145/1805986.1806000

[5] Bailey, C., and Pearson, E. 2010. An educational tool to support the accessibility evaluation process. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)* (W4A '10). ACM, New York, NY, USA. DOI=10.1145/1805986.1806003 http://doi.acm.org/10.1145/1805986.1806003

[6] Baguma, R., Stone, R. G., Lugega, J. T., and van der Weide, T. P. 2009. A framework for filtering web accessibility guidelines. In *Proceedings of the 2009 international Cross-Disciplinary Conference on Web Accessibililty (W4a)* (Madrid, Spain, April 20 - 21, 2009). W4A '09. ACM, New York, NY, 46-49. DOI= http://doi.acm.org/10.1145/1535654.1535663

[7] Brajnik, G. 2008. A comparative test of web accessibility evaluation methods. In *Proceedings of the 10th international ACM SIGACCESS Conference on Computers and Accessibility* (Halifax, Nova Scotia, Canada, October 13 - 15, 2008). Assets '08. ACM, New York, NY, 113-120. DOI= http://doi.acm.org/10.1145/1414471.1414494

[8] Brajnik, G. 2009. Validity and reliability of web accessibility guidelines. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility* (Assets '09). ACM, New York, NY, USA, 131-138. DOI=10.1145/1639642.1639666 http://doi.acm.org/10.1145/1639642.1639666

[9] Brajnik, G., Yesilada, Y. and Harper, S. 2010. Testability and validity of WCAG 2.0: the expertise effect. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility* (ASSETS '10). ACM, New York, NY, USA, 43-50. DOI=10.1145/1878803.1878813 http://doi.acm.org/10.1145/1878803.1878813

[10] Kelly B., Sloan D., Phipps L. Petrie H. and Hamilton F. 2005. Forcing Standardization or Accommodating Diversity? A Framework for Applying the WCAG in the Real World. *Proceedings of W4A at WWW2005: International Cross-Disciplinary Workshop on Web Accessibility*. New York: ACM Press.

[11] Kelly, B., Sloan, D., Brown, S., Seale, J., Petrie, H., Lauke, P and Ball, S. 2007. Accessibility 2.0: people, policies and processes. In *Proceedings of the 2007 international cross-disciplinary conference on Web accessibility (W4A)* (W4A '07). ACM, New York, NY, USA, 138-147. DOI=10.1145/1243441.1243471 http://doi.acm.org/10.1145/1243441.1243471

[12] Kelly, B., Nevile, L., Draffan, E., and Fanou, S. 2008. One world, one web .. but great diversity. In *Proceedings of the 2008 international Cross-Disciplinary Conference on Web Accessibility (W4a)* (Beijing, China, April 21 - 22, 2008). W4A '08, vol. 317. ACM, New York, NY, 141-147.

[13] Law, C., Jacko, J., and Edwards, P. 2005. Programmer-focused website accessibility evaluations. In *Proceedings of the 7th international ACM SIGACCESS Conference on Computers and Accessibility* (Baltimore, MD, USA, October 09 - 12, 2005). Assets '05. ACM, New York, NY, 20-27. DOI= http://doi.acm.org/10.1145/1090785.1090792

[14] Leporini, B., Paternò, F., Scorcia, A. 2006. Flexible tool support for accessibility evaluation, *Interacting with Computers*, v.18 n.5, p.869-890, September, 2006 DOI=10.1016/j.intcom.2006.03.001

[15] Lopes, R., Konstantinos, V., Carrico, L., Tzovaras, D. and Spiridon, L. 2010. The semantics of personalised web accessibility assessment. In *Proceedings of the 2010 ACM Symposium on Applied Computing* (SAC '10). ACM, New York, NY, USA, 1440-1441.

DOI=10.1145/1774088.1774394
http://doi.acm.org/10.1145/1774088.1774394

[16] Naftali, M. 2010. Analysis and integration of web accessibility metrics. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)* (W4A '10). ACM, New York, NY, USA. DOI=10.1145/1805986.1805996 http://doi.acm.org/10.1145/1805986.1805996

[17] Sloan, D., Gregor, P., Rowan, M,. and Booth, P. 2000. Accessible accessibility. In *Proceedings on the 2000 conference on Universal Usability* (CUU '00). ACM, New York, NY, USA, 96-101. DOI=10.1145/355460.355480 http://doi.acm.org/10.1145/355460.355480

[18] Sloan, D. 2002. Auditing accessibility of UK Higher Education web sites. *Interacting with computers 14 (2002).* 313-325.

[19] Sloan, D., Heath, A., Hamilton, F., Kelly, B., Petrie, H., and Phipps, L. 2006. Contextual web accessibility - maximizing the benefit of accessibility guidelines. In *Proceedings of the 2006 international Cross-Disciplinary Workshop on Web Accessibility (W4a): Building the Mobile Web: Rediscovering Accessibility?* (Edinburgh, U.K., May 22 - 22, 2006). W4A '06, vol. 134. ACM, New York, NY, 121-131.

[20] Sloan, D. 2006. The Effectiveness of the Web Accessibility Audit as a Motivational and Educational Tool in Inclusive Web Design. Ph.D. Thesis, University of Dundee, Scotland. June, 2006.

[21] Sullivan, T., and Mason, R. 2000. Barriers to use: usability and content accessibility on the Web's most popular sites. In *Proceedings on the 2000 conference on Universal Usability* (CUU '00). ACM, New York, NY, USA, 139-144. DOI=10.1145/355460.355549 http://doi.acm.org/10.1145/355460.355549

[22] Trewin, S., Cragun, B., Swart, C., Brezin, J and Richards, J. 2010. Accessibility challenges and tool features: an IBM Web developer perspective. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)* (W4A '10). ACM, New York, NY, USA. DOI=10.1145/1805986.1806029 http://doi.acm.org/10.1145/1805986.1806029

[23] Velleman, E., Meerveld, C., Strobbe, C., Koch, J., Velasco, C. A., Snaprud, M. and Nietzio, A. 2007. *D-WAB4 Unified Web Evaluation Methodology*. Web Accessibility Benchmarking Cluster. Retrieved 14th February 2010: http://www.wabcluster.org/uwem1_2/UWEM_1_2_CORE.pdf

[24] Vigo, M., Kobsa, A., Arrue, M., and Abascal, J. 2007. User-tailored web accessibility evaluations. In *Proceedings of the Eighteenth Conference on Hypertext and Hypermedia* (Manchester, UK, September 10 - 12, 2007). HT '07. ACM, New York, NY, 95-104. DOI= http://doi.acm.org/10.1145/1286240.1286267

[25] Waller, A., Hanson, V., and Sloan, D. 2009. Including accessibility within and beyond undergraduate computing courses. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility* (ASSETS '09). ACM, New York, NY, USA, 155-162. DOI=10.1145/1639642.1639670. http://doi.acm.org/10.1145/1639642.1639670

[26] WCAG 2.0 Assessment Tool. ACCESSIBLE Applications Design and Development Project. Accessed 14th February 2011: http://www.iti.gr/accessible/WCAG2.0_WebAssessmentTool_v2.0.zip

[27] Web Accessibility Toolbar 2.0. *The Paciello Group.* http://www.paciellogroup.com/resources/wat-ie-about.html. Accessed: 27/11/10.

[28] W3C. 1999. Web Content Accessibility Guidelines Version 1.0. Retrieved January 20th 2010: http://www.w3.org/TR/WCAG10/

[29] W3C. 2008. Web Content Accessibility Guidelines Version 2.0. Retrieved January 20th 2010: http://www.w3.org/TR/WCAG20/

[30] Yesilada, Y. Brajnik, G. and Harper, S. 2009. How much does expertise matter?: a barrier walkthrough study with experts and non-experts. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility* (Assets '09). ACM, New York, NY, USA, 203-210. DOI=10.1145/1639642.1639678 http://doi.acm.org/10.1145/1639642.1639678