Comparing website accessibility evaluation methods and learnings from usability evaluation methods

Tania Lang December 2003

Abstract

Different disabilities affect peoples' use of information and communication technologies. Consequently, there is a need to ensure that websites are accessible to people with disabilities. At the same time resources are required to help designers make websites more accessible. An assessment of usability evaluation methods (UEMs) demonstrates that there a significant alignment between the measures used in both usability and accessibility.. Furthermore, many of the benefits and drawbacks associated with UEMs can also be shown to apply to AEMs.

The aim of this paper was to determine the most effective accessibility evaluation method to determine web accessibility and concludes that a fully integrated approach (combining automated and manual evaluation and accessibility testing) is the best approach. In addition it is concluded that for organisations with time and cost constraints, a 'discount accessibility' method that combines automated and manual evaluation is the most appropriate evaluation method.

The merits of a study questioning the effectiveness of the Web Content Accessibility Guidelines (WCAG) are also discussed which lead to suggestions for further research.

1. Introduction

"*The power of the Web is in its universality. Access by everyone regardless of disability is an essential aspect*." - Tim Berners-Lee, director and inventor of the World Wide Web (W3C, 2003a)

Sullivan and Matson (2000, p. 139) note that from its beginning, the World Wide Web (WWW) "was conceived and implemented as a platform-neutral, device-independent means of accessing information". Despite this intention, a large percentage of websites today are inaccessible to users with disabilities, even though accessibility guidelines have been freely and widely available for over 3 years. The percentage of websites that are accessible is not necessarily increasing, and may even be decreasing with the proliferation of graphics and animation rich content over recent years.

In order to help organisations make their websites accessible, a number of methods and tools have been developed by researchers, practitioners and Information Technology (IT) companies. There are different views about the best evaluation method, yet there are relatively few studies that have been conducted comparing AEMs. A review of the small number of papers available suggests that comparisons are difficult because different methods measure different variables. For this reason the validity of the conclusions drawn in some studies may be questionable.

Whilst there are relatively few papers comparing accessibility evaluation methods, there are numerous papers discussing usability evaluation methods (UEMs). There are many similarities between UEMs and AEMs. For instance, accessibility testing with users generally appear to use similar measures as those used in usability testing such as task completion, satisfaction and efficiency (although user satisfaction is often excluded as a measure in accessibility testing). Automated and manual techniques generally evaluate a website's accessibility against the WCAG 1.0 (in a similar manner to heuristic evaluations or expert reviews which evaluate a website's usability against design guidelines). As result, it is likely that there are learnings from previous research comparing UEMs that may be applied to AEMs.

This raises a number of questions which this paper seeks to answer.

1) Is there a correlation between accessibility and usability evaluation methods?
2) What accessibility evaluation methods (AEMs) are available and what are the benefits and drawbacks for different methods?

3) Does full compliance with WCAG 1.0 actually mean that a website is fully accessible to people with disabilities?
4) Accordingly, what is most effective accessibility evaluation method to determine true web accessibility?

This paper will attempt to address these questions. The first section will discuss studies comparing Usability evaluation methods (UEMs) and possible learnings for accessibility evaluation methods (AEMs)

The second section will provide an overview of accessibility evaluation methods and their benefits and drawbacks.

Finally, the validity and usefulness of the W3C WCAG 1.0 will be assessed followed by a discussion of the most effective accessibility evaluation method or combination of techniques to determine true web accessibility.

2. Background

According to the Australian Bureau of Statistics (1998), approximately 19% of Australians have a disability. Different disabilities and the approximate percentage of each are as follows (Bourk, 2000):

Disability	Percentage of total disabled
Physical impairment	70%
Hearing impairment or degrees of deafness	14%
Intellectual or other mental impairment	9%
Vision impairment	4%
Psychiatric impairment	2%

These figures exclude many older people who don't consider themselves to have disability as well as a significant percentage of the population who have an injury or temporary disability.

Of greater significance to the field of web site and application design is the percentage of people (estimated at 10%) who have a disability that affects their use of information and communication technologies (ICT) (RNIB, 2000).

In addition to the number of people with disabilities that affect people's use of ICT, approximately 6.2 million Australians also have poor literacy and numeracy skills of which over a third (2.6 million) have very poor literacy and numeracy skills (often associated with learning disorders such as dyslexia) which affect their ability to complete tasks due to the considerable difficulty they experience with written materials (ABS 1996).

Hence, the real percentage of people who have some sort of disability (permanent or temporary) or learning disorder that affects their use of ICTs is likely to over 20% of the population.

2.1. How different disabilities affect peoples' use of ICT and implications for web and interface designers

The effects of different disabilities on web accessibility, the assistive technologies people use, and the implications for web and interface designers are well documented for many disabilities such as vision impairment. For instance, when designing for visually impaired users with screen readers, it is important to ensure there is ALT text on all images, the use of generic links such as "click here" is avoided (as when read to a user with a screen reader "Click here" would have no meaning).

Much less is widely known and understood about the effect of other disabilities on web accessibility such as cognitive impairment, motor skill impairment or learning disorders such as dyslexia.

Some guidelines do exist that encompass the needs of users with cognitive or neurological disabilities. For instance, designers need to ensure that:

- navigation is consistent and appropriate language is used; and
- flickering or strobing designs are avoided (W3C, 2003a).

However, guidelines associated with learning disorders such as dyslexia are generally only available from specific organisations dedicated to dyslexia and associated learning disorders. Common interface design problems that affect individuals with dyslexia when using computers and accessing websites are noted by Vassallo (2003) and include:

- small fonts;
- poor contrast backgrounds (either too low or too high);
- large blocks of text;
- cluttered pages;
- animated images or blinking/moving text;
- automated page or form redirects;
- lots of capitals or italics;
- fully justified text (resulting in uneven spacing between words); and
- wordy and confusing use of English.

Given the wide range and varying levels of different disabilities and the resulting implications of each on web and interface design, it is obvious, as noted in the literature, "there is a clear and immediate need for content providers to be aware of any accessibility problems currently affecting their resources, and how best to overcome these problems" (Sloan et al, 2002, p.314). However, it is often very difficult for designers and developers to fully comprehend and design accessible sites and this information has traditionally not been included in many IT and web development courses. Accordingly, the World Wide Web Consortium (W3C) established a Web Accessibility Initiative(WAI) to provide guidance for designers and developers to help make sites

accessible for people with disabilities. Details about W3C and the Web Accessibility Initiative are outlined in greater detail in Section 3.3.

2.2. Reasons for ensuring websites are accessible

As noted by W3C (2003a):

"*the Web is becoming a key resource for:*

- *news, information, commerce, entertainment,*
- *classroom education, distance learning,*
- *job searching, workplace interaction,*
- *community participation, government services.*

It is displacing traditional sources of information and interaction –

- *schools, libraries, print materials, discourse of the workplace;*
- *some of the traditional resources were accessible; some not".*

Accordingly, there is an increasing necessity to ensure that the World Wide Web offers unprecedented access to information for people with disabilities.

Broadly speaking, there are arguably three fundamental reasons why organisations should make their websites accessible:

Social reasons – organisations should feel obliged to be socially responsible and work towards removing discrimination and promoting human rights (HREOC, 2002). This entails making their website accessible to all users, including those with disabilities.

Economic / business reasons – it does not make economic sense for organisations to alienate 10% of potential users who may have a disability (HREOC, 2002). Having an inaccessible site may result in users with disabilities leaving for competitor sites or switching to more expensive channels, such as call centres and walk-in branches.

Legal reasons – In 1993 United Nations released its *Standard Rules on the Equalisation of Opportunities of Persons with Disabilities.* This document is only a guideline rather than international law and includes the recommendation that nations "develop strategies to make information services and documentation accessible for different groups of persons with disabilities" (Caslon Analytics, 2002). Many countries, such as Australia, had already developed such legislation prior to the release of the United Nations rules.

In Australia, people with disabilities are protected by the Disability Discrimination Act 1992 (DDA) which requires organisations to offer equal access for people with a disability in the area of information and online services, where it can reasonably be provided.

Specific guidelines are outlined by the Human Rights and Equal Opportunity Commission in the "World Wide Web Access: Disability Discrimination Act Advisory Notes Version 3.2" issued August 2002 (HREOC, 2002). These advisory notes are intended to assist organisations in developing or modifying accessible web pages thereby ensuring compliance with the DDA.

Organisations that fail to make their sites accessible in Australia run the risk of legal action, such as that taken against the Sydney Organising Committee for the Olympic Games (SOGOG) by Bruce Maguire (HREOC, 2000). In the case of Maguire vs SOGOG, Bruce Maguire lodged a complaint with HREOC under the DDA against SOCOG in June 1999. The basis for his complaint was that the SOCOG website was inaccessible to him as a blind person. On 24 August 2000, HREOC released its decision and supported Maguire's complaint, ordering certain accessibility changed be made to the Olympics.com site by mid September 2000. SOCOG failed to act on the ruling and were subsequently fined $20,000 (NuBlog, 2001).

2.3. W3C - resources to help make sites accessible

The World Wide Web Consortium, commonly referred to as 'W3C':

- *"is an international, vendor-neutral consortium, with over 400 Members*
- *promotes evolution & interoperability of the Web*
- *and has a strong focus on the universality of the Web"* (W3C, 2003a).

W3C has 4 "domains", one of which is the WAI initiative which is sponsored by governments and industry. The WAI initiative coordinates with the other 3 domains of the W3C and operates internationally. WAI has five areas of work:

1. *"ensuring that Web technologies support accessibility*
2. *developing guidelines for accessibility*
3. *improving tools to evaluate and repair web accessibility*
4. *developing materials for education and outreach*
5. *coordinating with research and development* (W3C, 2003a)".

2.3.1. W3C Web Content Accessibility Guidelines and evaluation methods

"Web Content Accessibility Guidelines 1.0" (WCAG 1.0) were developed by the "Web Accessibility Initiative" (WAI) to promote accessibility to designers and developers to encourage them to develop accessible web sites for people with a disability. The guidelines are intended to make web content accessible for all users regardless of the technology they use or their environment.

According to W3C (1999),

> *The guidelines are written for a variety of audiences - people who are designing Web sites; people who are checking existing Web sites for accessibility; organizations that wish*

to require a given level of accessibility for their Web sites; and others who are interested in ensuring that people with disabilities can access information on the Web.

They include 14 broad guidelines (65 specific guidelines) which are general principles of accessible design. Each guideline has a number of checkpoints describing how to apply that guideline to particular features of web pages. The core aim of the Web Content Accessibility Guidelines is to reduce the number of barriers on web pages for people with disabilities.

Further to the development of these guidelines, WAI have also developed a general approach for the evaluation of conformance to the Web Content Accessibility Guidelines (W3C, 2002). These are intended to be used during development and for evaluation of established websites. W3C (2002), also note that there are a number of other methods for evaluating web site accessibility and that no single method will capture all accessibility issues. The range of methods and their benefits and drawbacks will be discussed in greater detail in Section 4.4.

3. Discussion

This section is divided into three sub sections. The first section describes usability evaluation methods and reviews benefits, drawbacks and resulting learnings applicable to accessibility evaluation methods. The second section addresses the benefits and drawbacks of accessibility evaluation methods. The third section discusses the validity and usefulness of the W3C WCAG 1.0 and discusses the most effective accessibility evaluation method or combination of techniques to determine true web accessibility.

3.1. Comparison of usability evaluation methods (UEMs) and learnings for accessibility evaluation methods (AEMs)

Studies and discussions regarding AEMs tend to be relatively recent with most articles having been published since 2000. In contrast, as noted by Gray and Salzman (1998), usability evaluation methods have been of great interest to human-computer interaction (HCI) researchers and practitioners since the early 1990s and numerous studies have been conducted comparing the effectiveness of these methods (Jeffries et al, 1991; Karat et al, 1992; Nielsen, 1991; Desurvire et al, 1992; Nielsen & Phillips, 1993; Lindgaard, 1999; Wixon, 2003; Law & Hvannberg, 2002 and Doubleday et al, 1997). Of relevance to a comparison of AEMs are the criteria used to evaluate UEMs and the actual measures used.

3.2. Criteria used when comparing usability evaluation methods

A range of different criteria have been used in previous studies to compare usability evaluation techniques which include:

- ability to detect problems - number and severity of issues identified (Doubleday et al, 1997; Jeffries et al, 1991);
- accuracy and quality of the results (Doubleday et al, 1997);
- time, effort and cost effectiveness (Law & Hvannberg, 2002; Doubleday et al, 1997; Jeffries et al, 1991);
- usefulness of results and who can use the results – ability for developers to understand and implement required changes (Gray and Salzman, 1998; Doubleday et al, 1997; Wixon, 2003; Jeffries et al, 1991); and
- generality of results – extend to which results are representative of all users, which is often overlooked according to Gray and Salzman (1998).

Arguably all of these criteria can be applied to accessibility evaluation methods.

3.2.1. Measures used when evaluating usability and accessibility

Gray and Salzman (1998, p. 206) criticise the poor definition of measures used in early studies comparing UEMs and make the observation that "although something is being measured, it is far from obvious that these measures really reflect sensitivity to usability". Perhaps as a result of this widely acknowledged paper (which is referred to in many other articles), more recent studies have been more specific in outlining the specific measures they have used.

Typical measures used for each method are outlined in Table 1 along with the studies and evaluation methods documenting the use of these measures. A comparison of measures used for usability and accessibility evaluation clearly show there is significant alignment between UEMs and AEMs.

Table 1 - Comparison of evaluation approach and measures used for accessibility and usability.

Usability Evaluation		Accessibility Evaluation	
Measure	**Study / UEM(s)**	**Measure**	**Study / AEM(s)**
Number of usability problems and severity	• Sullivan & Matson (2000) / Automated (LIFT online) • Lindgaard (1999) / Heuristic evaluation • Law & Hvannberg (2002) / Heuristic evaluation • Doubleday et al (1997) / Heuristic evaluation	Number of accessibility problems / of W3C compliance issues and priority	• Sullivan & Matson (2000) / Automated and manual • Cooper & Rejmer (2001) / Automated (Bobby) • Sloan (2002) / Automated and manual • Zaphiris et al (2001) / Automated (Bobby)
Task completion	Usability testing • Law & Hvannberg (2002) / Usability testing	Task completion	• Sloan (2002) / Usability and accessibility testing • Colwell & Petrie (1999) / Accessibility testing
User satisfaction, opinions and emotional expression	• Lindgaard (1999) / Usability testing • Macleod et al (1997) / Usability testing • Law & Hvannberg (2002) / Usability testing • Doubleday et al (1997) / Usability testing	User satisfaction and opinions	• Sloan (2002) / Usability and accessibility testing • Colwell & Petrie (1999) / Accessibility testing
Duration – time to finish task	(2002) / Usability testing • Doubleday et al (1997) / Usability testing	Duration – time to finish task	• Sloan (2002) / Usability and accessibility testing

3.3. Learnings from studies undertaken comparing usability evaluation methods

Killam and Holland (2001) argue that the practices of usability evaluation apply to accessibility evaluation. Section 4.2.1 and in particular Table 1, clearly demonstrate there is significant alignment between the measures used in usability and accessibility evaluation methods. This raises the question, what learnings can be applied from previous studies so that we can build on the industry's existing knowledge base rather than unnecessarily replicating previous research?

One key insight is highlighted by Law and Hvannberg (2002) who note an emerging trend which is to delineate trade-offs – advantages and disadvantages of different UEMs – rather than strongly advocate a particular UEM. From their own research they note that different methods identify distinct sets of usability problems and that no one method highlights all problems. In their view, the different usability evaluation methods actually complement each other. Gray and Salzman (1998, p. 242) also advocate an approach which attempts to evaluate usability through "multiple converging measures". Applying these findings in relation to UEMs to AEMs, one could argue that no one method is appropriate and a combination of AEMs will yield better results.

Another key insight is in regards to the measures used to evaluate accessibility. It was noted at a workshop of accessibility experts (LaPlant et al, 2001), that accessibility evaluation methods often limit measures to efficiency and effectiveness whereas usability testing usually also includes a measure of user satisfaction (as highlighted in Table 1). The workshop consensus was that accessibility testing should measure efficiency, effectiveness and user satisfaction – the same measures typically used in usability evaluations.

Table 2 outlines the benefits and drawbacks of different UEMs as outlined in the literature: expert reviews (or "discount usability" methods) and usability testing. Many of these benefits and drawbacks apply to different AEMs as well. In particular, many of the benefits and drawbacks associated with expert reviews / heuristic evaluations are identical to the benefits and drawbacks associated with manual and automatic AEMs. Similarly, benefits and drawbacks associated with usability testing also are identical to those associated with accessibility testing.

One notable exception is the generality of results obtained from accessibility testing (which may be less representative of all users than results obtained from usability testing). This is because in accessibility testing, results may depend on the specific disability of the user and the specific version of assistive technology they use.

Table 2 – Outline of the benefits and drawbacks for two different usability evaluation methods.

Evaluation technique	Benefits	Drawbacks
Expert review ("discount usability" method or desktop evaluation using heuristic checklists, cognitive walkthroughs and/or guidelines)	• Not only identifies problems but also their underlying causes (Doubleday et al, 1997). • Much more cost effective e.g. one study took 9 hours for a heuristic evaluation compared with 200 hours for usability testing (Law & Hvannberg, 2002). Another study took 33.5 hours for a heuristic evaluation compared with 125 hours for usability testing (Doubleday et al, 1997).	• Evaluator online behaviour is different to user behaviour i.e. is not task based and accordingly, evaluators may miss major issues e.g. in one study, 39% of usability problems were missed by expert HCI evaluators (Doubleday et al, 1997). Another study claims that generally only 30-50% of end-user problem types are predicted (Law & Hvannberg, 2002). • Requires evaluator to have a greater skill level to review, understand and recommend solutions. • Limited availability of experienced evaluators (Law & Hvannberg, 2002) and general requirement to use more than one evaluator (Molich & Jeffries, 2003). • Evaluation tends to be more subjective – relying on the judgement of evaluators (Doubleday et al, 1997). • Problems identified often not distinct i.e. some problems may in fact be subsets of other problems (Doubleday et al, 1997). • May identify false positives – problems that don't actually affect users (Molich & Jeffries, 2003). • Less likely to identify positive usability features (Law & Hvannberg, 2002).
Usability testing (often referred to as user testing)	• More accurate and objective results (Law & Hvannberg, 2002) which are much more likely to be accepted and actioned by developers and managers. • Good at identifying problems experienced by real users actually using the system (Doubleday et al, 1997). Accordingly, generality of results will be greater than for heuristic evaluations. • More likely to identify major issues that may prevent users from completing tasks (Doubleday et al, 1997; Molich & Jeffries, 2003).	• Time consuming and expensive e.g. in one study, it took 125 hours for end user testing compared with only 33.5 hours for heuristic evaluation (Doubleday et al, 1997; Gray, 1995; Jeffries et al, 1991). • Highlights problems but does not identify the underlying cause or how to fix the problem (Doubleday et al, 1997) and as highlighted by Wixon (2003), problem detection is only the first step to improving a system. • Users tend to be more critical of themselves than the interface and post-testing questionnaires may be excessively favourable (Doubleday et al, 1997). • Will only identify problems actually encountered during tasks and specific areas of the site visited. Some problems may be missed accordingly and quality of the results is thus very dependent on the quality of the experimental tasks. (Doubleday et al, 1997). • Problems identified may not be related to the system but rather to design flaws with task scenarios e.g. poor wording (Law & Hvannberg, 2002). • Results may not be representative of all users if only one user experiences a problem – may be associated with the idiosyncrasies of individual test participants (Law & Hvannberg, 2002). • Often fails to identify minor problems (Jeffries et al, 1991).

3.4. Overview of accessibility evaluation methods and their benefits and drawbacks

A range of different accessibility evaluation methods exist but as highlighted by researchers such as (Sloan et al, 2002; Rowan et al, 2000) there are benefits and drawbacks with each method. The following section outlines the different evaluation techniques followed by an summary of their benefits and drawbacks as documented in the literature.

3.4.1. Automatic validation and evaluation tools

A large number of free and commercial automatic evaluation tools are now available and are well documented in the literature (Zaphiris et al, 2001). A large list of these tools are also outlined on the W3C website (http://www.w3.org/WAI/ER/existingtools.html#General).

Three of the more popular tools include:

- **Bobby** (http://bobby.watchfire.com/bobby/html/en/index.jsp) - a free evaluation tool that tests individual web pages against the guidelines established by the World Wide Web Consortium's (W3C) Web Access Initiative (WAI), as well as US Section 508 guidelines from the Architectural and Transportation Barriers Compliance Board. Results include an annotated image of the page being evaluated and a description of errors with links to detailed solutions (Cooper & Rejmer, 2001). Bobby is widely used. However, there are problems with the tool as highlighted by (Caslon Analytics, 2002). These researchers note that it is common to see Australian government and commercial sites with the 'BOBBY Approved' seal that still have accessibility problems.
- **Web Accessibility Visual Evaluator (WAVE)** (http://www.wave.webaim.org/index.jsp) – a free evaluation tool that performs automated checks and highlights potential issues requiring human judgement. Arguably much more usable and easier to decipher than Bobby results. Is the result of Pennsylvania's Initiative on Assistive Technology (PIAT) and developed at the Institute on Disabilities at Temple University and WebAIM (Web Accessibility in Mind) at the Center for Persons with Disabilities (CPD) at Utah State University.
- **UsableNet LIFT**(http://www.usablenet.com/products_services/products_services.html) range of commercial products that test and monitor for compliance with W3C and U.S. Section 508 accessibility guidelines and other usability guidelines

3.4.2. Manual evaluation with WCAG 1.0)

Manual evaluation with accessibility guidelines typically involves a number of steps (Rowan et al, 2000; W3C, 2002):

1. Examining a sample of pages (such as the home page, a form, a search page and a page with tables) using W3C's Web Content Accessibility Guidelines (WCAG) Checklist.
2. General inspection of the site under various settings and conditions such as:
   - with images turned off;
   - with sound turned off;
   - with font size increased (using browser settings);
   - with frames not loaded;
   - with style sheets not loaded;
   - with scripts, style sheets, and applets not loaded;
   - using lower screen resolutions (e.g. 800 X 600 which is still very common);
   - in grey scale;
   - using keyboard only; and
   - using different browsers.

An extensive manual evaluation generally also involves:

1. viewing a sample of pages with a text browser (such as Lynx) and a voice browser (such as JAWS) to establish if equivalent information and functionality is available and if information is presented in a meaningful order.
2. Reading a sample of pages to check if the text is clear and simple and appropriate for the audience of the site?

3.4.3. Accessibility testing with people with disabilities

Accessibility testing with people with disabilities typically involves the following (W3C, 2002):

- Testing the site using people with different disabilities, different levels of Internet experience, and different levels of familiarity with the site, using a variety of assistive technologies. May be conducted in a usability laboratory or in the test participant's own environment (which is often easier because they are used to their own system and adaptive technology).
- Depending on the testing goals, may involve giving participants tasks or scenarios to complete or may involve simply letting participants freely explore the site.
- Areas of the site that cause confusion or are difficult or impossible the use are noted.
- User satisfaction and opinions are captured at the end of the session through an interview or post-test questionnaire.

Table 3 – Outline of the benefits and drawbacks for different accessibility evaluation methods

Evaluation technique	Benefits	Drawbacks
Automatic validation and evaluation tools	• Some evaluation tools can guide developers through the repair of tables, images, scripts and links (UsableNet). • Evaluator can have lower skill level. • More cost effective. • Some evaluation tools can review a whole site e.g. check ALT text on all images across a site (Sloan et al, 2000; UsableNet). • Can evaluate a website in much less time than a human evaluator (Killam & Holland, 2001). • Minimise chances of missing accessibility issues (Killam & Holland, 2001).	• Cannot fully automatically check against all W3C guidelines e.g. accuracy and meaning of ALT tags - can only identify potential accessibility issues which need to be checked manually (Zaphiris et al, 2001; Kasday, 2000; Rowan et al, 2000). • Most tools (except for WAVE) cannot check if screen readers will read a page in a sensible order (Kasday, 2000). • None of the tools provide a set of easy to interpret results or specific and tailored expert recommendations (Sloan et al, 2000; Gibson et al, 2001; Rowan et al, 2000). • Sole reliance on automated tools may inaccurately determine that a website is accessible to a person with disabilities giving developers a false sense of security (Killam & Holland, 2001). Conversely, automatic tools can find a site inaccessible when in fact it may provide an acceptable level of accessibility (Rowan et al, 2000). • Some tools not internationalised and available in a range of languages and as such, may not be usable for non-English speaking evaluators (Cooper & Rejmer, 2001).
Manual evaluation against WCAG 1.0	• More likely to identify a wide range of accessibility problems for users with different disabilities • More cost effective.	• Requires evaluator to have a greater skill level to review, understand guidelines and recommend solutions. • Not practical to manually check a large number of pages (Sloan et al, 2000; Rowan et al, 2000). • Not likely to highlight usability issues which may prevent users with a disability (or even users without a disability) from completing their tasks (Killam & Holland, 2001).
Accessibility testing	• Arguably the most accurate technique for determining true accessibility of a site for users with specific disabilities and specific assistive technologies. Can identify problems that would be missed when checking against WCAG 1.0 (Colwell & Petrie, 1999; Killam & Holland, 2001). • More likely to uncover usability issues which may be applicable to all users, including those with no disability (Killam & Holland, 2001).	• Generality of results (Gray and Salzman, 1998; LaPlant et al, 2001) – may be limited depending on type and extent of test participants' disabilities and the specific assistive technologies they use. There is a danger in limiting focus on a specific disability e.g. visual impairment and a specific assistive technology e.g. JAWS screen reader, however it is also difficult to test representatives from a number of different disability groups. • Finding and obtaining access to test participants may be difficult (LaPlant et al, 2001). • Testing with people with disabilities can be a logistical challenge (LaPlant et al, 2001). • As per usability testing, accessibility testing can be time consuming and expensive. • May be difficult to determine if issues are accessibility issues or usability issues applicable to all users – need to test with people with no disabilities first to establish a baseline (LaPlant et al, 2001). Generally if an issue is only experienced by users with a disability, it can be categorised as an accessibility issue.

3.5. Validity and usefulness of the W3C WCAG 1.0

Whilst there has been significant industry involvement in the development of the WCAG 1.0, Colwell & Petrie (1999) question their effectiveness and efficiency in helping web authors develop accessible websites.

Colwell and Petrie (1999) conducted an experiment with 12 students who were learning HTML. Students were asked to adapt an existing web page making it accessible by following the WCAG 1.0. Results suggested that several improvements could be made to the WCAG 1.0 in order to achieve W3C's stated goal of helping people who are designing accessible web sites. These included improving: "their structure and tone; navigation within and between the documents; the content and presentation of examples; and additional information to be provided" (Colwell & Petrie, 1999, p. 12). It is worth noting that a second version of the WCAG is currently being developed and a working draft is available for comment. Hopefully, these issues have been addressed to some extent in this later version (W3C, 2003b).

Of perhaps greater significance was a second experiment conducted by Colwell and Petrie (1999) which involved accessibility testing of the pages developed by students using WCAG 1.0. 20 visually impaired participants used a wide variety of browsers and screen readers and performed the test in their own environment. Results showed that even though the web sites were designed using WCAG 1.0, some major usability issues still existed for some participants with specific browsers. In particular, 6 of the 15 participants could not view ALT text that was available (which appeared to be linked to browsers used and participant experience). Other results showed that some design changes made by the developers that were not based on the WCAG 1.0 actually appeared to improve accessibility more than some of the changes outlined in the guidelines (Colwell & Petrie, 1999).

Whilst there is insufficient literature available by which to assess Colwell and Petrie' findings, their experimental design and conclusions appear to be valid and worthy of further research and investigation. This is to ensure that organisations following the WCAG 1.0 do not develop a false sense of security and assume their website must be accessible for all users if they follow all the guidelines. In the meantime, the value of the WCAG 1.0 should not be underestimated as they play a valuable role in ensuring websites are accessible to people with disabilities.

3.6. Most effective accessibility evaluation method or combination of techniques to determine true web accessibility

It appears to be widely recognised in the literature that no one method can be effectively used to evaluate the accessibility of a website (Sloan et al, 2002; Sloan et al 2000; Kasday, 2000; Gibson et al, 2001, W3C, 2002; LaPlant et al; 2001; Rowan et al, 2000). This correlates with similar conclusions which have been reached in regards to usability evaluation methods (see section 4.3).

However, three evaluation methods - that often combine different individual methods - appear to be most favoured in the literature as follows:

1) '**Discount accessibility' methods** - these combine automatic and manual evaluation techniques and include the use of automated tools to check the site as well as a manual check (Winberg, 1999). This approach is arguably better than no accessibility review for organisations with cost and time constraints and should ensure websites offer a reasonable level of site accessibility.

2) **Accessibility testing with users** - Killam & Holland (2001) argue that the practices of usability apply and that whilst discount accessibility evaluation methods are cost effective and valuable (as they are for discount usability evaluation methods), accessibility testing with users is the only true test of success. LaPlant et al, 2001, argue that testing to guidelines without user testing may not ensure that a site is accessible for all users. However, whilst this method has significant benefits, the major drawbacks associated with using this method - being the generality of results (Gray and Salzman, 1998; LaPlant et al, 2001) and the cost – suggest that this method used in isolation appears to be inadequate.

3) **Fully integrated accessibility evaluation method** – This approach combines all of the above techniques and is supported by a number of experts (Rowan et al, 2000; Rosenbaum, 2001; Sloan et al, 2002). W3C refer to this type of approach as a "comprehensive evaluation [that] combines semi-automatic, manual and user testing of accessibility features" (2002). The following is a brief excerpt from the comprehensive evaluation process outlined by W3C:

   1 - Identify and disclose scope of site to be evaluated and the targeted conformance level for the evaluation …

   2 - Semi-automatic and automatic evaluation - Use at least two accessibility evaluation tools on page selection and run at least one tool across entire web site…

3 - *Manual evaluation - Examine page selection using relevant checkpoints from the Checklist of Checkpoints for Web Content Accessibility Guidelines 1.0… and examine page selection with graphical user interface (GUI) browsers: select at least three different configurations …*

4 - *Usability testing of accessibility features - Have people with different disabilities, different levels of technical expertise, and different levels of familiarity with the site, using a variety of assistive technologies and adaptive strategies, review page selection and explore freely across entire web site. Ask testers to try to find answers to the most common questions for which people visit the web site. Note areas where it is difficult or impossible the use the web site…*

5 - *Summarize and follow-up - Summarize any problems and best practices identified for each page type and a representative URL, and method by which they were identified Recommend follow-up steps*

If resources are available, a fully integrated approach is arguably the best approach and is most likely to determine the greatest number of accessibility issues as well as achieve a high level of accessibility, usability and user satisfaction for people with disabilities. However this approach may not be appropriate for organisations with time and cost constraints which is often the case.

In environments with time and cost constraints, a 'discount accessibility' method which combines automated and manual evaluation is arguably the most appropriate evaluation method and studies outlined in this paper suggest that this method is significantly better than individual methods used in isolation (such as fully automated evaluation).

4. Further research required

As discussed, automated and manual accessibility evaluations generally evaluate the accessibility of web sites by determining the conformance level against WCAG 1.0. However, accessibility testing with real users generally appears to use measures such as task completion, satisfaction and efficiency. Accordingly it is difficult to compare evaluation techniques as they actually measure different variables.

The research problem that this literature review initially aimed to address was to compare and assess accessibility evaluation tools and techniques and determine the most effective accessibility evaluation method or combination of techniques.

However, as Killam and Holland (2001) state, the best evaluation of accessibility is "ultimately a question of the end user's ability to locate information or exercise functionality that determines accessibility". Hence, the question proposed for further research is as follows:

> **Does full compliance with W3C WCAG 1.0 actually mean that a website is accessible to people with different disabilities using different assistive technologies and if not, what is the best technique for determining if a site is truly accessible?**

This question worthy of further investigation to ensure that organisations following the WCAG do not develop a false sense of security and possibly falsely assume their website must be accessible for all users if they follow all the guidelines.

5. Conclusion

Given the significant percentage of people who have a disability that affects their use of ICT; there are three important reasons for ensuring websites are accessible, namely:

- social reasons;
- economic / business reasons; and
- legal reasons.

Despite these reasons, many organisations still have sites that are not accessible to people with disabilities.

The W3C WAI along with accessibility researchers, practitioners and IT companies have developed a range of guidelines and resources to help organisations ensure their websites are accessible. Yet some controversy exists regarding the best AEM.

Previous studies comparing UEMs demonstrate there is significant alignment between the measures used in usability and accessibility evaluation methods and the criteria used to compare evaluation methods.

Learnings that may be applied when comparing AEMs include the general consensus that no one evaluation method alone can identify all problems. Furthermore, many of the benefits and drawbacks associated with UEMs can also be show to apply to AEMs.

Following a review of the literature and the different methods, it can be argued that if resources are available, a fully integrated approach (combining automated and manual reviews as well as accessibility testing) is arguably the most effective approach and most likely to determine the greatest number of accessibility issues and achieve a high level of accessibility, usability and user satisfaction for people with disabilities. However, for organisations with time and cost constraints, a 'discount accessibility' method which combines

automated and manual evaluation is arguably the most appropriate, cost effective evaluation method.

Finally, as a result of this literature review, the following research question is proposed and warrants further investigation:

> Does full compliance with W3C WCAG 1.0 actually mean that a website is accessible to people with different disabilities using different assistive technologies and if not, what is the best technique for determining if a site is truly accessible?

6. References

ABS (1998) ABS survey shows one Australian in five has a disability. Retrieved September 23, 2003, from http://www.abs.gov.au/ausstats/abs@.nsf/5e3ac7411e37881aca2568b0007afd16/81a21b2987d05d49ca2568a90013631f!OpenDocument&Highlight=0,disability

Bourk, M. (2000). Telecommunications Policy In Australia and People with Disabilities. Paper presented at Internet World 2000, 8 June 2000, Sydney. Retrieved September 23, 2003, from http://www.tomw.net.au/uso/w3d.html

Caslon Analytics. (2002). Caslon Analytics Online Accessibility Guide. Retrieved November 7, 2003, from http://www.caslon.com.au/accessibilityguide2.htm

Colwell, C. & Petrie, H. (1999). Evaluation of Guidelines for Designing Accessible Web Content. Position paper [Electronic version]. *INTERACT'99 Workshop: Making Designers Aware of Existing Guidelines for Accessibility.*

Cooper, M. & Rejmer, P. (2001). Case study: localization of an accessibility evaluation [Electronic version. *Proceedings of the ACM CHI2001 Conference on Human Factors in Computer Systems.* Seattle: ACM.

Deshpande, Y., Chandrarathna, A., & Ginige, A. (2002). Web site auditing: first step towards re-engineering [Electronic version]. *Workshop on web engineering* at *Proceedings of the 14th International Conference on Software Engineering and Knowledge Engineering.* Ischia: Italy.

Desurvire, H.W., Kondziela, J.M. & Atwood, M.E. (1992). What is gained and lost when using evaluation methods other than empirical testing [Electronic version]. *Proceedings of the HCI'92 Conference on People and Computers, VII.* New York: Cambridge University Press.

Doubleday, A., Ryan, M., Springett, M. & Sutcliffe, A. (1997). A comparison of usability techniques for evaluating design [Electronic version]. *Proceedings of the conference on Designing interactive systems: processes, practices, methods, and techniques*. Amsterdam

Gibson, L., Sloan, D., Gregor, P. & Booth, P. (2001). Auditing for accessibility [Electronic version]. *Paper presented at The Tenth International World Wide Web Conference*. Hong Kong.

Gray, W. (1995). Discount or disservice?: discount usability analysis - evaluation at a bargain price or simply damaged merchandise? (panel session) [Electronic version]. *Conference on Human Factors and Computing Systems*. Denver: ACM.

HREOC, (2002). World Wide Web Access: Disability Discrimination Act Advisory Notes Version 3.2. Retrieved 3 November, 2003, from http://www.hreoc.gov.au/disability_rights/standards/www_3/www_3.html

HREOC, (2000). Bruce Lindsay Maguire v Sydney Organising Committee for the Olympic Games. Retrieved September 23, 2003, from http://www.hreoc.gov.au/disability_rights/decisions/comdec/2000/DD000120.htm

Jeffries, R., Miller, J.R., Wharton, C. & Uyeda, K.M. (1991). User interface evaluation in the real world: A comparison of four techniques [Electronic version]. *Proceedings of the ACM CHI'91 Conference on Human Factors in Computer Systems.* New York: ACM.

Karat, C.M., Campbell, R. & Fiegel, T. (1992). Comparison of empirical testing and walkthrough methods in user interface evaluation [Electronic version]. *Proceedings of the ACM CHI'92 Conference on Human Factors in Computer Systems.* New York: ACM.

Kasday, L.R. (2000). A tool to evaluate universal Web accessibility [Electronic version]. *Proceedings of the CUU 2000 ACM Conference on Universal Usability.* Arlington, VA.

Killam, B. & Holland, B. (2001) Position Paper on the Suitability to Task of Automated Utilities for Testing Web Accessibility Compliance [Electronic version]. *The Usability SIG Newsletter: Usability Interface Accessibility and Usability: Partners in Effective Design. April 2003.*

Kurniawan, S.H. & Zaphiris, P. (2001). Usability and Accessibility Comparison of Governmental, Organizational, Educational and Commercial Aging/Health Related Web Sites [Electronic version]. *Paper presented at Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomics Society.* Minneapolis, MN.

LaPlant, W.P., Laskowski, S.J. & Stimson, M.J., 2001. Report on UPA Workshop 6 Exploring Measurement and Evaluation Methods for Accessibility [Electronic version]. *Workshop at Tenth*

Annual Usability Professionals' Association Conference.

Law, L., Hvannberg, E.T. (2002). Complementarity and convergence of heuristic evaluation and usability test: a case study of universal brokerage platform [Electronic version]. *Proceedings of the Second Nordic Conference on Human-computer Interaction, October 19-23, 2002.* Denmark: Aarhus.

Lindgaard, G. (1999). Of babies, bath water and usability techniques in web evaluations [Electronic version]. *OZCHI '99 Interfaces for the Global Community.* Wagga Wagga.

Macleod, M., Bowden. R., Bevan, N. & Curson, I. (1997). The MUSiC performance measurement method [Electronic version]. *Behaviour and Information Technology 16*, 1-27.

Molich, R. & Jeffries, R. (2003). Comparative Expert Reviews [Electronic version]. *CHI2003: New Horizons.* Ft. Lauderdale: ACM.

Nielsen, J. (1992). Finding usability problems through heuristic evaluation [Electronic version]. *Proceedings of the ACM CHI'92 Conference on Human Factors in Computer Systems.* New York: ACM.

Nielsen, J. & Phillips, V.L. (1993). Estimating the relative usability of two interfaces: Heuristic, formal, and empirical methods compared [Electronic version]. *Proceedings of the ACM INTERCHI'93 Conference on Human Factors in Computer Systems. New York: ACM.*

NuBlog (2001). Reader's guide to Sydney Olympics accessibility complaint. Retrieved November 3, 2003, from http://www.contenu.nu/socog.html

Rosenbaum, S. (2001). Exploring Measurements and Evaluation Methods for Accessibility [Electronic version]. *Paper presented for UPA 2001 Workshop.* Lake Las Vegas, Nevada.

Rowan, M., Gregor, P., Sloan, D. & Booth, P. (2000). Evaluating web resources for disability access [Electronic version]. *The Fourth International ACM Conference on Assistive Technologies (ASSETS 2000)*. Virginia: ACM.

RNIB (2000). *Websites that work*. [Videotape]. United Kingdom: Royal National Institute of the Blind.

Sloan, D., Gibson, L., Booth, P. & Gibson, L. (2002). Auditing accessibility of UK Higher Education web sites [Electronic version]. *Interacting with Computers 14*, 313-325.

Sloan, D., Gregor, P., Rowan, M. & Booth, P. (2000). Accessible accessibility [Electronic version]. *Proceedings of the CUU 2000 First ACM Conference on Universal Usability*. Arlington, VA: ACM.

Sullivan, T. & Matson, R. (2000). Barriers to use: usability and content accessibility on the Web's most popular sites [Electronic version]. *Proceedings of the CUU 2000 ACM Conference on Universal Usability.* Arlington, VA: ACM.

UsableNet (2003). LIFT Products and Services. Retrieved November 3, 2003, from http://www.usablenet.com/products_services/lift_dw/lift_dw.html

Vassallo, S. (2003). Enabling the Internet for People with Dyslexia. Retrieved November 3, 2003, from http://e-bility.com/articles/dyslexia.shtml

W3C. (1999). Fact Sheet for "Web Content Accessibility Guidelines 1.0". Retrieved September 23, 2003, from http://www.w3.org/1999/05/WCAG-REC-fact

W3C. (2001). How People with Disabilities Use the Web. Retrieved September 23, 2003, from http://www.w3.org/WAI/EO/Drafts/PWD-Use-Web/#constnav

W3C. (2002). Evaluating Web Sites for Accessibility. Retrieved November 3, 2003, from http://www.w3.org/WAI/eval/

W3C. (2003a). Overview of the Web Accessibility Initiative. Retrieved September 23, 2003, from http://www.w3.org/WAI/

W3C. (2003b). Web Content Accessibility Guidelines 2.0 W3C Working Draft 24 June 2003. Retrieved November 3, 2003, from http://www.w3.org/TR/WCAG20/

Winberg, F. (1999). Discount accessibility engineering: Haven't we met before? [Electronic version]. *INTERACT'99 Workshop: Making Designers Aware of Existing Guidelines for Accessibilty.* Edinburgh: UK

Wixon, D. (2003). Evaluating Usability Methods [Electronic version]. *Interactions July August 2003.*

Zaphiris, P., Kurniawan, S.H. & Ellis, R.D. (2001) Usability and Accessibility Comparison of Governmental, Organizational, Educational and Commercial Aging/Health-Related Web Sites [Electronic version]. *WebNet Journal: Internet Technologies, Applications & Issues 3* (3), 45-52.