Understanding and Supporting Web Developers:

Working Practices and Resources for the Creation and Evaluation of Accessible Websites

David Mark Swallow

Doctor of Philosophy

University of York

Computer Science

November 2017

Abstract

Web developers have a responsibility to develop websites that are accessible and usable by the broadest range of users, including people with disabilities. Despite numerous initiatives over the last two decades to support, encourage and compel developers to fulfil this responsibility, websites continue to exhibit persistent accessibility problems. This programme of research aimed to better understand the difficulties that developers face and to develop the necessary support for them to successfully integrate accessibility into their existing workflows. The first study systematically reviewed 397 web accessibility evaluation studies published over a 15-year period between 1999 and 2014. This showed a persistent occurrence of accessibility problems that does not appear to be improving. The second study followed a contextual design methodology to investigate the working practices of 13 professional developers. This showed how they are hindered, not by limited awareness or concern, but by a lack of knowledge and practical guidance on how to make websites accessible. To understand the nature of their confusion and uncertainty, the third study interviewed 26 professional developers and attempted to elicit their mental models of web accessibility. Their mental models were found to incorporate some, but not nearly enough, knowledge and awareness of accessibility and were based on a conceptual model that prioritises technical conformance over user experience. These findings were embodied in the design and implementation of an accessibility information resource, called WebAIR. The ease of use and effectiveness of WebAIR in supporting the creation and evaluation of accessible websites was evaluated in a series of four studies with both professional and novice developers under increasingly realistic experimental conditions. The resource was well received by participants in each study and, despite concerns over its viability within organisations that place little value on web accessibility, WebAIR was demonstrated to be a usable, pragmatic accessibility information resource.

Table of Contents

List of Tables

List of Figures

Acknowledgements

A number of major life events have occurred since I started this PhD research in 2011. In 2013, I married my wife, Sarah. Later that year, we moved house. The following year, 2014, our first daughter, Olivia, was born. Two years later, in 2016, I changed jobs, leaving my role of Research Associate at the University of York to work as an Accessibility Engineer for The Paciello Group. Two years after that, in 2018, our second daughter, Sophie, was born. Remaining focused on my research during this eventful period of my life, not to mention keeping up with the latest developments in the ever-changing field of web accessibility, has been no mean feat and I could not have achieved it without the tireless commitment and support of many people.

I am extremely grateful to all of the participants who took part in this research for their valuable input and assistance. Being fortunate enough to have found many talented professional web developers willing to give up their time has allowed me to ensure their profession is accurately represented and adequately supported by this research.

I thank my supervisors, Prof. Helen Petrie and Dr. Christopher Power for their longstanding support and guidance, both as colleagues and mentors, during my time at the University of York.

I thank the many friends, colleagues and office mates with whom I have worked over the years: Anna Bramwell-Dicks, Lucy Buykx, Alistair Edwards, André Freire, Andrew Lewis, John Precious, and Tanja Walsh. In particular, I thank Andreas Savva for offering me a sounding board for my ideas as well as many laughs and much encouragement.

I owe a huge thank you to my parents, Joan and Kenn, and to my in-laws, Tony and Carol, for their continued love and support (and also for cutting the grass and tending the garden while I've been holed up inside working on my thesis).

Above all, I would like to express my sincere gratitude to my wife, Sarah, and daughters, Olivia and Sophie, without whose unconditional love, steadfast support and limitless patience none of this would have been possible, or worth it.

Declarations

I declare that this thesis is a presentation of original work and I am the sole author, unless otherwise indicated in the text. This work has not previously been presented for an award at this, or any other, university. All sources are acknowledged as references. Parts of this thesis have been used in the following publications:

Swallow, D., Petrie, H., & Power, C. (2016). Understanding and Supporting Web Developers: Design and Evaluation of a Web Accessibility Information Resource (WebAIR). In *Proceedings of the 3rd International Conference on Universal Design (UD 2016)* (pp. 246-255). Amsterdam, The Netherlands: IOS Press.

Swallow, D. (2015). Improving Accessibility Support for Web Developers. In *Adjunct Proceedings of the 15th IFIP TC. 13 International Conference on Human-Computer Interaction (INTERACT 2015)* (Vol. 22, p. 55). Bamberg, Germany: University of Bamberg Press.

Swallow, D., Power, C., Petrie, H., Bramwell-Dicks, A., Buykx, L., Velasco, C. A., Parr, A., & Connor, J. O. (2014). Speaking the Language of Web Developers: Evaluation of a Web Accessibility Information Resource (WebAIR). In Miesenberger K., Fels D., Archambault D., Peňáz P., and Zagler W. (Eds.) *Computers Helping People with Special Needs. ICCHP 2014. Lecture Notes in Computer Science, vol 8547* (pp. 348-355). Berlin/Heidelberg, Germany: Springer.

Petrie, H. (Ed.) (2013). i2Web Deliverable 4.8: Field studies of the use of i2Web services and applications.

Petrie, H. (Ed.) (2013). i2Web Deliverable 4.7: Final user-based evaluations of accessibility and usability of services and applications.

Petrie, H., Power, C., & Swallow, D. (Eds.) (2011). i2Web Deliverable 3.2: Requirements for web developers and web commissioners in ubiquitous Web 2.0 design and development.

Petrie, H., Power, C., Swallow, D., Velasco, C.A., Gallagher, B. Magennis, M., Murphy, E., Collin, S., & Down, K. (2011). The value chain for web accessibility: challenges and opportunities. In Weber, G., Petrie, H., and Darzentas, J. (Eds.) *Proceedings of Accessible Design in the Digital World (ADDW 2011).*

The first part of the research presented in this thesis (which includes a contextual inquiry of web developer working practices, and the design, implementation and initial evaluation of a web accessibility information resource) was conducted under the auspices of the i2Web (Inclusive Future-Internet Web Services) project[1], on which I worked as a Research Associate at the University of York.

The aim of the i2Web project was to develop tools for transforming websites to meet people's needs and to produce tools to help web developers, web accessibility experts and web commissioners to deliver accessible websites. The project ran from November 2010 to April 2013 and was co-financed by the 7th Framework Programme of the European Commission under its ICT Programme (Grant Agreement number 257623).

The remaining research (which includes further development and evaluation of the accessibility information resource) was self-funded and conducted entirely on my own.

[1] http://i2web.eu/

Chapter 1. Introduction

1.1. This is for everyone

During the opening ceremony of the London 2012 Olympic Games—a spectacular celebration of British history and culture—a mock-up of a typical suburban house rose into the air to reveal a man sitting at a computer. That man was Sir Tim Berners-Lee, computer scientist and inventor of the World Wide Web. At that moment, he sent a tweet[2] that read: "This is for everyone". The message was immediately spelt out in LCD panels attached to the chairs of the 80,000 audience members, broadcast to millions of television viewers around the world, and re-tweeted more than 10,000 times online (see Figure 1.1). By "this", Berners-Lee was referring to the web, and by "everyone", he meant all people, regardless of race, gender, age, income, religion, or disability.

Figure 1.1: Screenshot of tweet sent by Sir Tim Berners-Lee during the opening ceremony of the London 2012 Olympic Games.

The World Wide Web (from here on referred to as 'the web') is an important resource that is having a growing and profound influence upon all aspects of society, including education, health care, access to government, commerce, employment, recreation and many more. To realise Berners-Lee's vision of a global hyperlinked information system that is free and open to everyone, it is vital that all members of society can access it. This is particularly the case for people with disabilities, for whom the web offers the potential for unprecedented access to information (Henry, 2005a). The United Nations (UN) Convention on the Rights of Persons with Disabilities (2006) recognises such access as a basic human right. The practice of ensuring that people of all abilities and disabilities can access the web regardless of the technology they use and the method in which they access it is called "web accessibility" (Paciello, 2000).

2 https://twitter.com/timberners_lee/status/228960085672599552

1.2. Diversity in disability

Web accessibility is a complex, diverse and multi-faceted practice that has borne a range of definitions. In lieu of a widely agreed and used definition, Petrie, Savva and Power (2015) synthesised 50 definitions of web accessibility, drawn from a range of books, papers, standards, guidelines and online sources. Their lengthy but comprehensive unified definition, which this thesis adopts, defines web accessibility as,

> all people, particularly disabled and older people, can use websites in a range of contexts of use, including mainstream and assistive technologies; to achieve this, websites need to be designed and developed to support usability across these contexts. (p. 3)

Implicit in the first part of this definition is an acknowledgement of the diversity of people who benefit from accessible websites. Like any other group of web users, disabled people are a diverse population with differing needs and requirements. Users may experience, for instance, varying degrees of visual, auditory, physical, cognitive, and age-related disabilities that could hinder their access to the web (Thatcher et al., 2006). To mitigate these barriers, the design and development of websites must incorporate the needs and requirements of different user groups across different use contexts. While everyone's experience of the web need not be identical, websites should provide a comparable user experience for disabled people (Henry, Abou-Zahra, & White, 2010).

The technologies that disabled people use to access the web are equally diverse. In addition to mainstream devices and web browsers, some disabled people use adaptive hardware and software tools, known as assistive technologies (ATs), to display and interact with websites. Someone with visual disabilities, for example, may use a screen reader to read aloud the contents of webpages, or a screen magnifier to enlarge areas of the screen. Someone with physical disabilities that prevent them from using a keyboard and mouse may use alternative input devices (such as gaze/eye tracking equipment or speech recognition software). Although manufacturers have made considerable efforts to standardise how browsers and other user agents render websites, there are still considerable differences between them[3] that must be taken into consideration.

A substantial set of accessibility recommendations, techniques and best practices has evolved to support the diversity of disabled web users and the technologies they use.

[3] http://www.html5accessibility.com

Some recommendations involve the provision of alternative or enhanced content, such as textual descriptions of images; audio description of video; and captioning, transcription, or sign language translation of audio. Some techniques focus on coding websites so as not to interfere with or impede user agents and ATs, such as marking up and structuring the code to make sure webpages are navigable and operable using keyboards and alternative input devices. Some best practices relate to the planning, writing and managing of content, such as providing clear labels and instructions and avoiding unnecessarily complex words and phrases. Underscoring each recommendation, technique or best practice is an emphasis on clear communication and flexibility of content – a key principle of web accessibility (Henry, 2005a).

The diversity of approaches to creating accessible websites is mirrored by a variety of methods and tools for evaluating them. Automated tools may be used for examining website code to identify accessibility issues. However, due to limitations in the type of issues that such tools can programmatically determine (that is, test without human intervention), manual inspection methods are also necessary. These methods rely on the judgement of human evaluators, who must examine websites and their underlying code for potential accessibility issues. Additional evaluation may also be carried out with disabled users to gain a more thorough understanding of a website's accessibility issues. User testing methods involve observing people using websites to complete representative tasks. Each method has its strengths and weaknesses and no single method is effective in identifying all accessibility issues (see Section 2.1). A combination of automated evaluation, manual inspection and, preferably, user evaluation is generally considered necessary to robustly determine web accessibility.

Current approaches to both creating and evaluating accessible websites tend to focus on conformance to a set of accessibility guidelines. However, as explored further in Section 2.5.3, certain limitations of existing guidelines threaten the effectiveness of these approaches. An empirical study by Power, Freire, Petrie and Swallow (2012) demonstrates that accessibility guidelines either fail to cover or inadequately address many of the actual problems encountered by blind web users. Power et al. argue it is time to move away from the problem-based paradigm for web accessibility towards a more design principle-based approach. This corresponds to the second part of the unified definition of web accessibility by Petrie et al. (2015), which places emphasis on designing and developing websites to support usability across different contexts of use.

1.3. The value chain of accessible web development

In addition to the individuals, teams, or organisations that commission, own and manage websites ("web commissioners"), the value chain of accessible web development typically involves numerous internal and external stakeholders (see Figure 1.2). In large enterprises and small and medium enterprises (SMEs), internal stakeholders may comprise: web and graphic designers who contribute graphic or stylistic input; content providers and copy editors who create the content that populates websites; user experience (UX), usability and accessibility experts, who offer guidance and support; and quality assurance (QA) teams who test and audit websites. External stakeholders include operating system (OS), web browser, and AT providers who create the hardware and software platforms on which websites run; and web and accessibility standards groups who define the standards to which websites should conform. The attitudes, perceptions and actions of each of these stakeholders are crucial to the accessibility of websites for disabled people.

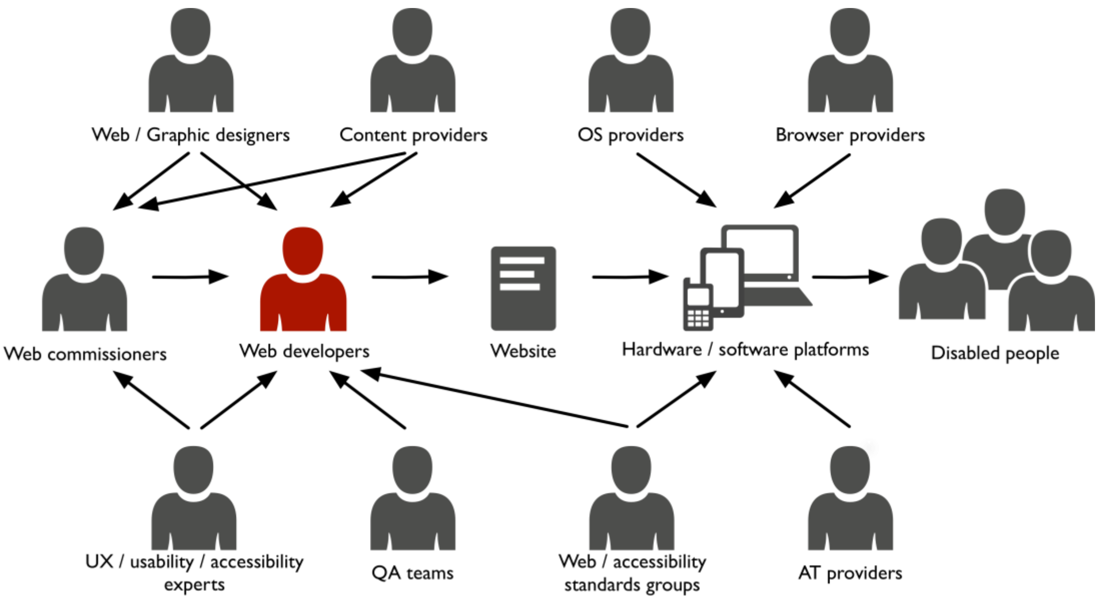


Figure 1.2: Stakeholders involved in the value chain of accessible web development

Amidst the various stakeholders, the individuals whom I would argue have the *greatest* responsibility for web accessibility are web developers, who must mediate and integrate input from numerous other stakeholders in the actual creation and evaluation of websites. I base this assertion on my direct experience of working as a professional web developer in a SME between 2004 and 2006. I am not alone, however, in making this assertion: respondents in surveys of web developers (e.g. Lazar, Dudley-Sponaugle, & Greenidge, 2004; Yesilada, Brajnik, Vigo, & Harper, 2015) typically consider themselves responsible, or, at least, *the most* responsible for web accessibility. Moreover, the web developers I encountered during this programme of research frequently expressed ultimate responsibility for creating accessible websites.

It is hard to precisely define the role of web developer (the variety of job titles and responsibilities reported by participants in this research being testament to this: see Section 4.2.2) but for the purpose of this thesis, it is anyone who is engaged in programming (or coding) programs and web applications. Ultimately, web developers are responsible for writing and evaluating code that is supported by browsers, devices and ATs to provide access to information and functionality and to deliver an accessible user experience. This ideally requires an awareness of the diverse needs, requirements, capabilities and limitations of users; a familiarity with the variety and intricacies of browsers, devices and ATs that provide access to the web; and an understanding of the necessary coding techniques, design principles and programming practices to put accessibility into practice.

1.4. Guiding accessible web development

Since almost as long as the web has existed, there have been numerous initiatives to support, encourage and compel web developers to fulfil their responsibility to develop accessible websites. These initiatives include projects, working groups, and task forces, such as the World Wide Web Consortium's (W3C) Web Accessibility Initiative (WAI)[4], the WAI's Education & Outreach Working Group (EOWG)[5], and the (now defunct) Web Standards Project's (WaSP) Accessibility Task Force (ATF)[6]. Other initiatives include accessibility legislation, such as Australia's Disability Discrimination Act of 1992, the UK's Equality Act of 2010, and the US's Americans with Disabilities Act (ADA) of 1990. These efforts have resulted in a well-established body of accessibility information, often presented in the form of a set of guidelines or recommendations, such as the US Government's Section 508 standards and the WAI's Web Content Accessibility Guidelines (WCAG) (Chisholm, Vanderheiden, & Jacobs, 1999; Caldwell, Cooper, Reid, & Vanderheiden, 2008a).

[4] https://www.w3.org/WAI/
[5] https://www.w3.org/WAI/EO/
[6] https://www.webstandards.org/action/atf/index.html

Despite these well-intentioned initiatives and the wide availability of accessibility information, web developers are still failing to create accessible websites. This is apparent from the growing number of lawsuits[7], particularly in the US, filed against businesses due to inaccessible websites. Such lawsuits are filed to ensure that businesses provide their goods and services to people with disabilities without discrimination. The poor state of web accessibility is also evident from regular empirical studies that monitor the accessibility of websites from around the world and which frequently report substantial and repeated failings. Later in this programme of research, I systematically review 397 web accessibility evaluation studies published over a 15-year period between 1999 and 2014. This shows a persistent occurrence of accessibility problems that does not appear to be improving (see Chapter 3).

Existing attempts to explain the poor state of web accessibility rarely attribute it to a lack of awareness or concern on the part of web developers (e.g. Farrelly, 2011; Lazar et al., 2004). Numerous surveys of web developers over the last two decades, in fact, indicate a growing awareness, both of web accessibility and of associated guidelines, tools and legislation. Studies instead tend to cite societal and organisational factors, such as a lack of training and education in web accessibility (e.g. Freire, Russo, & Fortes, 2008) or difficulties in enforcing accessibility legislation (Olalere & Lazar, 2011). Some studies (e.g. Putnam et al., 2012) call attention to negligent client and organisational attitudes to web accessibility, whereas others (e.g. Brajnik, 2009; Petrie, King, Velasco, Gappa, & Nordbrock, 2007) highlight difficulties in using accessibility tools, guidelines and resources. The latter is particularly challenging for *novice* web developers, who have limited experience of creating websites, let alone making them accessible. Few, if any, studies, however, examine web developers in-depth in the context of their working environments to actually determine why and how they struggle to create accessible websites.

1.5. Research aims and objectives

The programme of research presented in this thesis aims to better understand the difficulties that experienced and novice web developers face in creating and maintaining accessible websites. In addition, it aims to contribute to the support necessary to enable web developers to better integrate web accessibility into their existing workflows. A

[7] http://www.karlgroves.com/2011/11/15/list-of-web-accessibility-related-litigation-and-settlements/

further aim is to investigate the effectiveness of this support in a series of evaluations conducted under laboratory and real-world conditions. To achieve these aims, this programme of research is divided into three broad stages: analysis, intervention, and evaluation.

The first stage—**analysis**—examines the current state of web accessibility and draws partly upon ethnographic methods to explore the current working practices of professional web developers. The first study within this stage is a systematic review of the web accessibility evaluation literature. The objective of this study is to investigate whether the accessibility of websites has improved over the last two decades. This is followed by a contextual inquiry investigation with 13 professional web developers. The objective of this second study is to examine web developers in context to gain a greater understanding of their current working practices and to establish how web accessibility may be integrated into their existing workflows. This detailed exploration is complemented by an interview study with 26 professional web developers. The goal of this study is to elicit web developers' mental models—the internal cognitive frameworks that represent their understanding of web accessibility—and to identify any knowledge gaps, misconceptions or errors that may account for their difficulties in making websites accessible. This stage addresses the following research questions:

- Has the accessibility of websites improved over the last two decades?
- How do web developers' current working practices support or hinder the creation of accessible websites?
- How do web developers' existing knowledge and understanding of web accessibility support or hinder the creation of accessible websites?

The second stage—**intervention**—builds on the research findings of the previous stage by focusing on the development and initial evaluation of a prototype accessibility information resource. The aim of this resource is both to support web developers in the creation and evaluation of accessible websites and also to foster a greater understanding of web accessibility. This stage documents a number of design decisions made in the development of the resource and presents an initial evaluation study with 26 professional web developers and 7 computer science students, who may be classed as novice web developers. The objective of this study is to gather web developers' initial impressions of the resource and to investigate the impact of the design decisions on web developers' attitudes, both towards the resource and towards web accessibility in general. The section concludes with an account of further refinements made to the resource in response to the gathered feedback. This stage addresses the following research question:

- What design attributes of an accessibility information resource are appropriate to support web developers in (a) the creation and evaluation of accessible websites; and (b) to foster a greater knowledge and understanding of web accessibility?

The third stage—**evaluation**—aims to evaluate the ease of use and effectiveness of the previously developed accessibility information resource in supporting web developers to create and evaluate accessible websites. The first study in this stage is a laboratory-based evaluation of the resource with 48 computer science students, who, again, may be classed as novice web developers. The objective of this study is to evaluate the effectiveness of the resource in supporting novice web developers to confidently identify accessibility problems and solutions in existing websites. This is followed by a real-world evaluation of the resource with 32 professional web developers. The goal of this study is to evaluate the effectiveness of each resource in supporting experienced web developers to identify accessibility problems and solutions in a purpose-built website. The final study is an "in the wild" evaluation of the resource with 15 professional web developers. The objective of this study is to gain a more detailed understanding of how experienced web developers use the resource and whether and how it can be integrated into their existing working practices to maintain accessibility support over time. This stage addresses the following questions:

- How effective is the resource in supporting novice web developers to confidently identify accessibility problems and solutions in existing websites?
- How effective is the resource in supporting professional web developers to identify accessibility problems and solutions in a purpose-built website?
- How effective is the resource in supporting professional web developers to create an accessible website?
- How easily can the resource be integrated into professional web developers' existing working practices?
- Does exposure to the resource have any impact upon web developers' knowledge and understanding of web accessibility?

My ultimate objective in undertaking this programme of research is to foster a greater understanding of web accessibility among web developers. I hope that my engagement with web developers through these studies will generate in them enthusiasm and passion for web accessibility, stimulate their interest in learning more about the needs and requirements of disabled users, and encourage them to recognise their role in developing a web that is for everyone.

1.6. Research contributions

The main contributions of this research will include:

- a systematic review of 397 web accessibility evaluation studies published over a 15-year period between 1999 and 2014;
- an empirical demonstration that web developers are hindered, not by limited awareness or concern, but by a lack of knowledge and practical guidance on how to make websites accessible;
- an empirical elicitation of web developers' mental models of web accessibility, which are revealed to be very limited and based on a faulty conceptual model that prioritises technical requirements over user needs and standards conformance over user experience;
- a definition of the necessary characteristics and attributes of accessibility information resources to support web developers in the creation and evaluation of accessible websites;
- an embodiment of these findings in the development of a prototype accessibility information resource, called WebAIR; and
- a comprehensive evaluation of WebAIR that demonstrates its ease of use and effectiveness in supporting web developers to create and evaluate accessible websites.

1.7. Thesis overview

This thesis is organised into 10 chapters and several appendices, containing material used in the various studies. The remaining chapters are organised as follows:

- Chapter 2 presents a detailed review of the literature. It begins with a comprehensive history of web accessibility initiatives that have been developed over the last two decades. This is followed by an exploration of the impact of these initiatives upon the attitudes and awareness towards web accessibility among web developers. It concludes with an examination of various influences upon the accessibility of websites and, specifically, why web developers struggle with existing accessibility tools, guidelines and resources.

- Chapter 3 describes a systematic review of web accessibility evaluation studies published over a 15-year period between 1999 and 2014.
- Chapter 4 describes a contextual inquiry investigation with 13 professional web developers.
- Chapter 5 explores the concept of mental models, which emerges as a possible means of representing web developers' understanding of web accessibility. The chapter presents a review of the mental models literature and describes an interview study with 26 professional web developers.
- Chapter 6 describes the development of an accessibility information resource to support web developers in the creation and evaluation of accessible websites. It also presents an initial evaluation of the resource with 26 professional web developers and 7 novice web developers.
- Chapter 7 describes an evaluation of the ease of use and effectiveness of the resource in supporting 48 novice web developers to confidently evaluate accessible websites.
- Chapter 8 describes an evaluation of the ease of use and effectiveness of the resource in supporting 32 professional web developers to evaluate accessible websites.
- Chapter 9 describes an "in the wild" evaluation of the ease of use and effectiveness of the resource in supporting 15 professional web developers to create accessible websites.
- Chapter 10 concludes the programme of research, providing a summary and discussion of the results, the contributions of the thesis, limitations and reflections on the work, as well as the scope for future research. It also reviews a number of other accessibility information resources that have been developed since this programme of research began.

Chapter 2. Literature review

2.1. Methods for determining web accessibility

To determine the accessibility of websites, evaluators (be they web developers, accessibility experts or other stakeholders) utilise a variety of approaches, including automated evaluation, manual inspection, and user evaluation of web accessibility.

2.1.1. Automated accessibility evaluation

Automated evaluation involves the use of tools that automatically examine webpage code to identify (and in some cases, propose solutions to) accessibility issues. Such tools evaluate the conformance of individual webpages or, sometimes, entire websites to a specific set of guidelines or requirements, such as WCAG, Section 508 or other similar principles. Automated tools can check the validity of HTML markup and Cascading Style Sheets (CSS), detect the presence or absence of certain features (such as labels and headings), and compare detected values (such as the colour contrast ratio between foreground and background colours) to values found in guidelines and specifications.

The outcome of these tests is then used to generate a report detailing where the issues lie in the webpage or website and what the web developer should do to address them. The format of the report differs from tool to tool. Some tools, such as AChecker[8] by the Inclusive Design Research Centre, provide text-based reports that describe specific issues or guidelines that have been violated (see Figure 2.1). Other tools, such as WAVE[9] by WebAIM, generate graphical annotations representing accessibility errors and warnings that are dynamically inserted into to the webpage code, allowing potential accessibility issues to be viewed in situ (see Figure 2.2). Still others, such as Testo Accesibilidad Web (TAW)[10] by the Spanish Fundación CTIC (Centre for the Development of Information and Communication Technologies in Asturias) generate reports using Evaluation and Reporting Language (EARL) (Abou-Zahra, 2007), which is a working draft published by the W3C for describing accessibility test results in a machine-readable (XML/RDF) format. Using EARL facilitates the comparison and aggregation of accessibility test results.

---

[8] https://achecker.ca

[9] http://wave.webaim.org

[10] http://www.tawdis.net

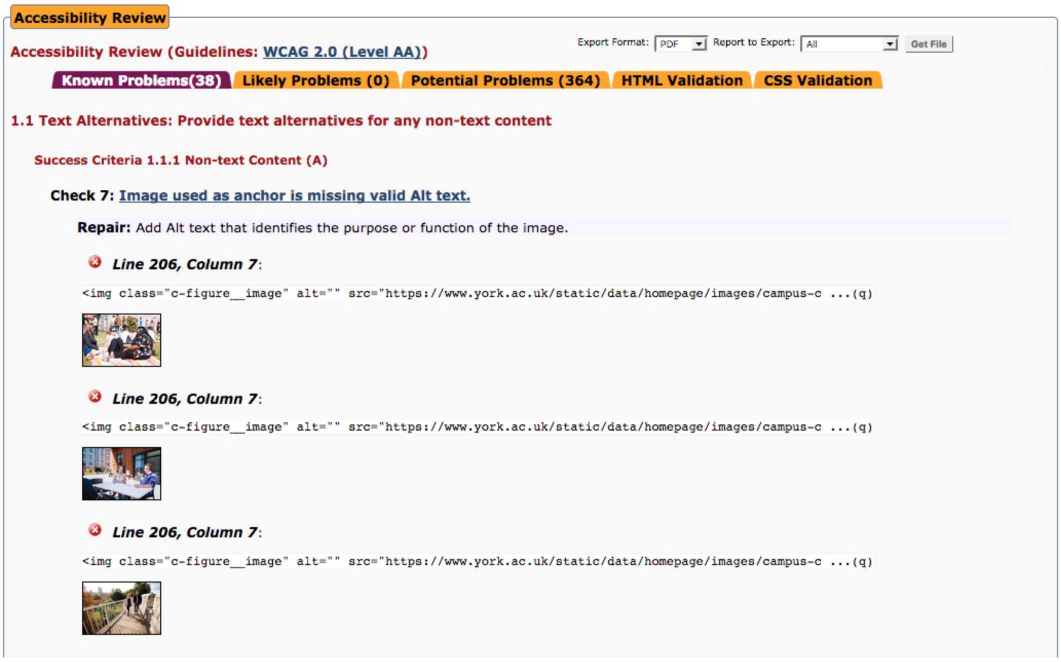


Figure 2.1: Example of report generated by the automated accessibility evaluation tool, AChecker by the Inclusive Design Research Centre. Reprinted under the GNU Lesser Public Licence, v2.1.

One of the earliest and most well-known automated evaluation tools was Bobby, which was originally developed by the Center for Applied Special Technology (CAST). Bobby was a free, public service for automatically determining conformance to a limited subset of WCAG 1.0 and Section 508, with successful websites being endorsed as "Bobby Approved". The Watchfire Corporation acquired Bobby in July 2002 and later incorporated it into their existing tool, WebXACT, in May 2005. Following IBM's acquisition of Watchfire in 2007, Bobby/WebXACT ceased to be available as a free service and was taken offline in February 2008[11]. Other popular examples of automated evaluation tools include, Cynthia Says[12] by Cryptzone, Tenon[13] by Karl Groves, and aXe[14] by Deque Systems. The WAI maintains an extensive list of over 90 web accessibility evaluation tools, around two-thirds of which can be used to automatically determine the accessibility of websites (Eggert & Abou-Zahra, 2016).

Automated evaluation tools typically report some metric of web accessibility. For conformance-based evaluation, this is usually a dichotomous pass/fail measurement of whether or not a webpage conforms to a particular set of guidelines or requirements.

---

[11] https://www.webmasterworld.com/accessibility_usability/3579037.htm
[12] http://www.cynthiasays.com
[13] https://tenon.io
[14] https://www.deque.com/products/axe

For example, if a webpage satisfies each of the Section 508 requirements (that a tool can determine automatically), it will be deemed to have passed Section 508 (subject to certain manual checks). WCAG 1.0 and 2.0 both provide an additional degree of differentiation by categorising individual guidelines according to three discrete conformance levels: Level A (lowest), AA, and AAA (highest). These reflect the impact of different guidelines upon users and provide an approximate measure or benchmark of a website's accessibility.

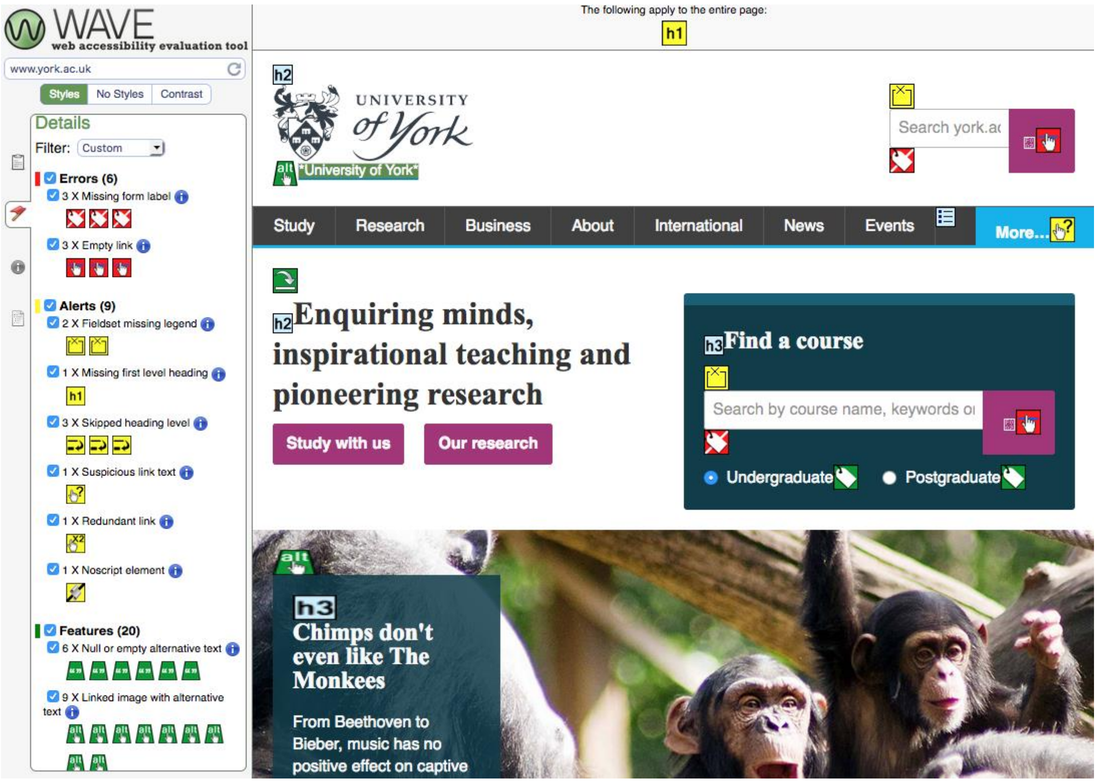


Figure 2.2: Example of report generated by the automated accessibility evaluation tool, WAVE by WebAIM. Reprinted with permission.

Due to the potential lack of sensitivity and precision afforded by the dichotomous nature of conformance, whereby entire webpages can fail due to a single violation, some authors (e.g. Parmanto & Zang, 2005; Sullivan & Matson, 2000; Vigo & Brajnik, 2011) consider it too blunt a measure. Instead, they have proposed more precise metrics, such as the failure rate metric (Sullivan & Matson, 2000), the WAB (Web Accessibility Barriers) metric (Parmanto & Zang, 2005), and the WAQM (Web Accessibility Quantitative Metric) (Vigo, Arrue, Brajnik, Lomuscio, & Abascal, 2007). These are intended to more precisely indicate the degree of accessibility by taking multiple factors into account, such as the number of potential failure points in a website, the number of guideline violations, and the severity of accessibility barriers.

Automated evaluation tools expedite the otherwise time-consuming and often tedious task of determining the accessibility of websites (Power, Freire, & Petrie, 2010). The tests they conduct are easily repeatable, allowing accessibility conformance to be compared over time. The tools can also be more easily integrated into the development process, increasing the likelihood of accessibility issues being remediated. However, due to limitations in the type of issues that automated tools can programmatically determine, they can only be used to evaluate websites against a limited subset of guidelines. For example, WCAG 2.0 Success criterion 1.1.1 states:

> **1.1.1 Non-text Content:** All non-text content that is presented to the user has a text alternative that serves the equivalent purpose, except for the situations listed below. (Level A)

To satisfy this criterion, all non-text content (such as images) must have an appropriate text alternative that that conveys the meaning, or purpose of the image. This is achieved using the `alt` attribute of the HTML `img` element (see Figure 2.3).

Figure 2.3: Image of the University of York logo. Reprinted with permission.

The image is marked up using the following syntax:
****

While tools can automatically determine the presence or absence of alternative text of images (by identifying the `alt` attribute) they cannot establish whether the text is sufficiently descriptive or even accurate. An evaluation of six state-of-the-art automated evaluation tools by Vigo, Brown and Conway (2013) estimates that they cover, at most, only 50% of the WCAG 2.0 success criteria and catch, at best, 38% of violations. Furthermore, due to apparent differences in how automated evaluation tools interpret accessibility guidelines, no two tools will necessarily return the same results (Harper & Chen, 2012). Also, despite their automation, such tools still require evaluators or web developers to interpret and verify what can often be overwhelmingly long and detailed results (Sloan, Gregor, Booth, & Gibson, 2002) in order to remediate any problems.

2.1.2. Manual accessibility inspection

Given the limitations of automated evaluation tools, and the number of guidelines that cannot be checked automatically, some degree of manual inspection is often necessary. Manual inspection relies on the judgement of human evaluators, who must examine webpages and their underlying code for potential accessibility issues. As with automated accessibility evaluation, manual inspection is often used to evaluate a website's conformance to a particular set of guidelines or requirements.

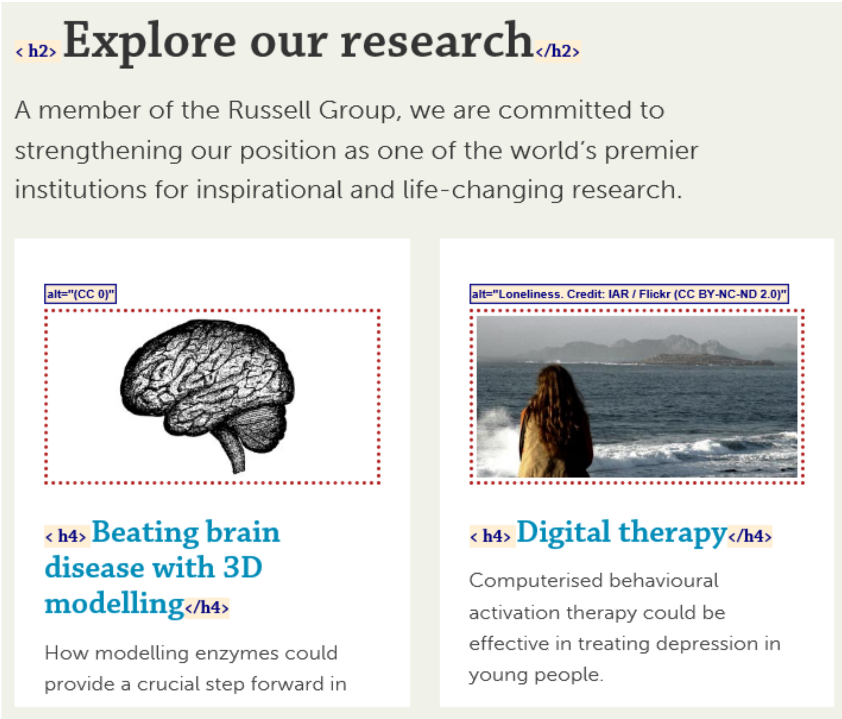


Figure 2.4: Output of the Web Accessibility Toolbar by The Paciello Group, highlighting webpage headings and displaying the alternative text of images. Reprinted with permission under the GNU Public Licence, v3.0.

Various tools and utilities typically assist evaluators using manual inspection methods. These include multi-purpose web browser toolbars, add-ons and extensions, such as the Web Accessibility Toolbar[15] by The Paciello Group that enable evaluators to interact with webpages in various ways, for example, viewing the alternative text of images, or highlighting headings, lists, tables and other structural components (see Figure 2.4). Some utilities enable evaluators to carry out very specific tasks, such as examining the colour contrast of text against its background (see Figure 2.5), adjusting the size of text, or determining the flicker rate of images or animation.

[15] https://developer.paciellogroup.com/resources/wat/

Manual inspection methods usually also involve viewing webpages across different web browsers and devices under different configurations (e.g. screen resolution, text size, window size, input mode, operating system etc.) to reflect a broad user experience. In addition, evaluators may use ATs, such as screen readers, screen magnifiers, and keyboard interfaces, to emulate the experiences of disabled users. For instance, this might involve turning off the display while using a screen reader to interact with a webpage, or using software that emulates how people with colour vision deficiencies might view a webpage. While evaluators cannot fully emulate the experience of disabled users by using ATs in this way, such methods can go some way to identifying potential accessibility issues for certain user groups.

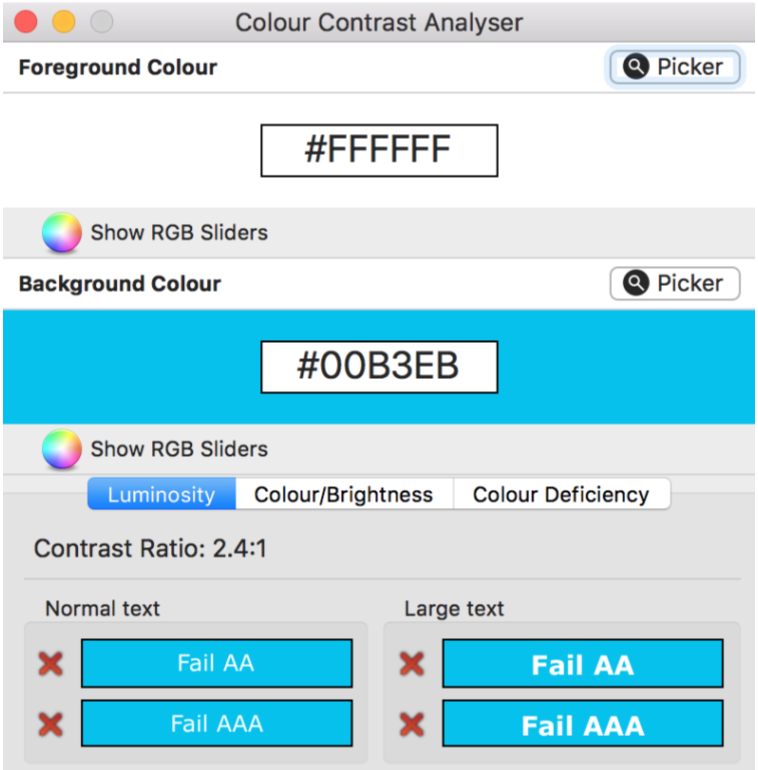


Figure 2.5 Output of the Colour Contrast Analyser by The Paciello Group, displaying the colour contrast ratio of text. Reprinted with permission under the GNU Public Licence, v3.0.

Some manual inspection methods focus less on evaluating the conformance of websites to a set of guidelines and more on identifying the specific problems that disabled users are likely to encounter. For instance, barrier walkthrough (Brajnik, 2006), which is an adaptation of a usability inspection method called heuristic walkthrough, involves one or more expert evaluators systematically "walking through" a website to identify potential accessibility barriers. Brajnik (2006) defines a barrier as "any condition that hinders the user's progress towards the achievement of a goal" (p. 3). Application of the barrier walkthrough method requires evaluators to first define the relevant user categories (e.g. blind users of screen readers or cognitively disabled users) and the relevant goals and scenarios to be considered. Evaluators then analyse selected webpages for a set of pre-determined barriers that are relevant to these user categories, goals and scenarios. Each barrier is also assigned a severity level. While initial evaluations of the barrier walkthrough method (Brajnik 2006, 2008) found it to be more precise than conformance testing (i.e. the reported problems were more likely to be true problems) and to report fewer false positives, they also found it to be less reliable, with independent evaluators tending to produce different results (Brajnik, 2008).

The disparity observed between evaluators using the barrier walkthrough method is symptomatic of the flaws in manual inspection on the whole. Many authors (e.g. Jaeger, 2006; Mankoff, Fait, & Tran, 2005; Olalere & Lazar, 2011) consider manual inspection a much more thorough and accurate approach to determining the accessibility of websites. It allows evaluators to identify a broad spectrum of accessibility issues, many of which cannot be identified programmatically. However, manual inspection is also very subjective and reliant upon the evaluator's knowledge, understanding and interpretation. This makes it hard to reliably repeat the tests (Martínez, Martínez-Normand, & Olsen, 2009). For example, two evaluators examining the image used in the earlier example (see Figure 2.3) may differ in their understanding of what constitutes sufficiently descriptive alternative text.

Manual inspection methods are also very time-consuming and do not scale easily to large numbers of webpages. Consequently, such methods are typically used in combination with automated evaluation. This combined approach was recommended in the Unified Web Evaluation Methodology (UWEM) (Velleman, Velasco, Snaprud, & Burger, 2006), which was the result of a combined effort of three European projects that made up the Web Accessibility Benchmarking Cluster. In lieu of official guidance from the W3C, the UWEM proposes a complete methodology for evaluating the conformance of websites to WCAG 1.0 and recommends statistical methods for sampling content as well as interpreting and reporting test results. In 2014, the W3C published their own Website Accessibility Conformance Evaluation Methodology (WCAG-EM) (Velleman & Abou-Zahra, 2014), which outlines a similar approach for evaluating the conformance of websites to WCAG 2.0. Although WCAG-EM does not mandate any particular evaluation method, it recommends the use of automated tools and manual inspection methods, as well as the involvement of users.

2.1.3. User evaluation of web accessibility

While a combination of automated evaluation and manual inspection can identify a broad spectrum of potential accessibility issues, an approach that yields a more comprehensive understanding of the accessibility of websites is user evaluation (Henry, 2010). Both automated evaluation and manual inspection tend to focus on evaluating conformance to accessibility guidelines such as WCAG or Section 508. While this takes into consideration a broad cross-section of users, it cannot match the richness of data offered by user testing. Also, while conformance evaluation identifies *potential* accessibility issues in websites, user testing reveals *actual* problems that users encounter.

User evaluation involves observing a group of people with different disabilities and different levels of expertise (both in using computers as well as using any necessary ATs) using websites to undertake a set of representative tasks. The participants' actions and comments are observed and recorded by the evaluator with the aim of identifying the accessibility issues that they may encounter.

A study by Freire (2012) investigated the relationship between user-based and guideline-based accessibility evaluations. This involved task-based user evaluations of 16 websites by a panel of 64 disabled users (32 blind, 19 partially sighted and 13 dyslexic) and manual audits of website conformance to WCAG 1.0 and 2.0. Comparisons between the two types of evaluation showed that as well as failing to address many user problems, the guidelines were often ineffective in avoiding them. These conclusions, which also confirmed previous findings by the UK Disability Rights Commission (2004) and Rømen and Svanæs (2008, 2012), underscore the need to involve disabled users in the design and evaluation of websites and highlight the danger in relying solely upon technical accessibility conformance.

Testing with users is widely regarded as the "gold standard" of accessibility evaluation (Aizpurua, Arrue, Harper, and Vigo, 2014). It goes beyond conformance evaluation to offer an insight into how people actually use websites and the accessibility barriers they may encounter. However, user testing requires considerable effort and resources, not only to recruit and compensate participants, but also to conduct the testing and analyse the results. For instance, it is difficult to recruit a representative sample of participants with different disabilities. Similarly, finding a large enough sample of participants to represent the broad selection of ATs available (as well as differing levels of expertise in using them) can prove challenging. Like manual inspection, user evaluation depends largely on the abilities of the evaluator, who must unobtrusively elicit and record accessibility issues without influencing the participant's behaviour. Also, user evaluation can generate a considerable amount of data that can be time-consuming to analyse.

Each of these approaches has its strengths and weaknesses and no single approach is effective in identifying all accessibility issues (Sloan et al., 2002; Klein et al., 2003; Jaeger, 2006). With this in mind, a combined methodology of automated evaluation, manual inspection and, preferably, user evaluation is generally recommended to robustly determine the accessibility of websites.

2.2. A history of web accessibility initiatives

For almost as long as the web has existed, there have been numerous initiatives to support, encourage and compel web developers and other stakeholders to fulfil their responsibility to develop accessible websites. One of the earliest initiatives was a set of guidelines developed by Gregg Vanderheiden at the Trace R&D Center at the University of Wisconsin, USA in January 1995 (Vanderheiden, 1995). Vanderheiden recognised a trio of components that contribute to the provision of accessible web content: the "viewer" (e.g. web browsers), the "pipeline" (e.g. web servers), and the "source material" (e.g. HTML markup). Acknowledging that web developers have little influence upon the accessibility of the former two, Vanderheiden instead focused upon the last, the HTML markup. The guidelines that Vanderheiden developed referred heavily to the rendering of HTML markup in Mosaic, an early graphical web browser. Despite this specificity, the guidelines provided general accessibility guidance on how to make HTML documents accessible to people with disabilities, given the technological possibilities at the time.

Vanderheiden's guidelines were somewhat cautionary, advocating the use of established HTML over more experimental features, and offering both immediate and potential future solutions to accessibility problems. This uncertainty may be attributed to the rapid evolution of the HTML specification at the time, which was undergoing numerous revisions by competing working groups. However, development of the HTML specification was unified in 1996, when maintenance of it was transferred to the W3C. The W3C is the main international organisation for the web, founded by Sir Tim Berners-Lee in 1994. Under the auspices of the W3C, the HTML specification soon stabilised, and in December 1997, version 4.0 of the HTML specification was published as a W3C Recommendation (Raggett, Le Hors, & Jacobs, 1997). Notably, HTML 4.0 was also the first HTML version to refer to accessibility in its preamble.

The increasing stability of the HTML specification resulted in a large number of accessibility guidelines being published by researchers from both industry and academia (e.g. Flavell, 1996; IBM, 1996; both cited in Vanderheiden & Chisholm, 1998). These guidelines ranged from advice on specific aspects of HTML (e.g. alternative text recommendations or table formatting) to broad design considerations. The original guidelines by Vanderheiden also evolved to reflect the continued development of the HTML specification and its growing implementation by browser manufacturers (Vanderheiden & Chisholm, 1996). In an attempt to unify the various sources of accessibility information, the numerous guidelines were brought together by the Trace R&D Center to form the Unified Web Site Accessibility Guidelines (UWSAG) (Vanderheiden & Chisholm, 1998).

The eighth and final version of UWSAG incorporated 38 different accessibility guidelines and overview documents. Control of the guidelines transferred to the W3C in 1998. Following their earlier adoption of the HTML specification, the W3C launched the WAI. The WAI brought together individuals and organisations from around the world with the aim of developing strategies, guidelines and resources to foster understanding and implementation of web accessibility (Henry & Brewer, 2000). The WAI also chartered various Working and Interest Groups to investigate different aspects of web accessibility and provide a forum for discussing web accessibility issues. Of particular relevance to web developers was the Education & Outreach Working Group (EOWG), which was chartered with developing training and education materials aimed at raising awareness of web accessibility (Brewer, 1998). Following UWSAG's transfer to the W3C, it became the foundation for the WAI's first publication, the Web Content Accessibility Guidelines (WCAG 1.0), which became an official W3C Recommendation in May 1999 (Chisholm et al., 1999). Like its predecessors, WCAG was a comprehensive set of guidelines aimed primarily at web developers that explained how to make web content accessible to people with disabilities.

WCAG 1.0 comprised 14 guidelines, representing general principles of accessible web design. Within each guideline were a number of checkpoints (65 in total). Each checkpoint described how web developers could adapt their web content to make it accessible and provided technology-specific guidance that they could use to validate conformance. To reflect its impact on the accessibility of content to different groups of disabled users, each checkpoint was assigned a priority level (Thatcher et al., 2006). The levels ranged from Priority 1 (the most basic level of web accessibility) through Priority 3 (the highest and most complex level of web accessibility). There were 16 Priority 1 checkpoints, 30 Priority 2 checkpoints, and 19 Priority 3 checkpoints. A webpage was said to be Level A conformant if it satisfied all Priority 1 checkpoints, Level AA conformant if it satisfied all Priority 1 and 2 checkpoints, and Level AAA conformant if it satisfied all Priority 1, 2 and 3 checkpoints. A supporting document, Techniques for Web Content Accessibility Guidelines 1.0, was published as a W3C Note the following year (Chisholm, Vanderheiden, & Jacobs, 2000).

Prior to the publication of WCAG 1.0, the US government drew upon the work of the WAI and other organisations to inform its own set of accessibility guidelines. In 1998, the United States Congress signed into law an amendment to the Workforce Rehabilitation Act of 1973. The new legislation, known as Section 508, required electronic and information technology (including websites) developed or purchased by federal agencies to be accessible to people with disabilities. The part of Section 508 that related specifically to web accessibility was Subpart B § 1194.22 (Web-based Intranet and Internet Information and Applications), which comprised 16 requirements, labelled a-p. Many of these requirements mapped directly or closely to WCAG 1.0 Priority 1 checkpoints (Thatcher et al., 2006). Whereas WCAG 1.0 conformance was voluntary, however, Section 508 became a legally enforceable set of accessibility requirements (albeit only applicable to websites of US federal agencies) in June 2001. Consequently, it emerged as an important benchmark for measuring web accessibility conformance in the US.

Recognition of the growing prominence of the web, coupled with increasing awareness of the challenges that people with disabilities encounter in accessing information, led to many countries and jurisdictions developing their own accessibility legislation and policies, including Australia, Austria, Belgium, Brazil, Canada, Denmark, the European Union (EU), Finland, France, Germany, Hong Kong, India, Ireland, Italy, Japan, Korea, Luxembourg, The Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Thailand and the UK (Thatcher et al., 2006; Mueller, Jolly, & Eggert, 2017).

Many countries adopted the WCAG 1.0 or Section 508 requirements in their entirety and incorporated them into legislation, local standards, and sector-specific guidelines.

To support legislation enforcement and encourage standards conformance, industrial and academic researchers developed a variety of web accessibility evaluation tools (see Section 2.1.1 and 2.1.2). Such tools could either generate an automated report of accessibility conformance or support web developers to carry out a manual assessment themselves. While the relative simplicity of such tools made them effective in promoting and encouraging web accessibility, their ability to test only a limited subset of guidelines frequently lulled web developers into a false sense of security.

Acknowledging that web developers were not solely responsible for creating an accessible user experience, the WAI developed two further sets of web accessibility guidelines, aimed at other stakeholders in the web accessibility value chain. The Authoring Tool Accessibility Guidelines (ATAG 1.0), which became a W3C recommendation in February 2000, acknowledged that web developers use authoring tools (e.g. code editors, content management systems, and blogging tools) to create web content. ATAG provided guidance to manufacturers not only on how to support web developers to create accessible web content but also on how to make authoring tools themselves accessible, so as to enable people with disabilities to create web content (Treviranus, McCathieNevile, Jacobs, & Richards, 2000). The User Agent Accessibility Guidelines (UAAG 1.0), which became a W3C recommendation in December 2002, acknowledged that users rely upon user agents (such as browsers, readers and media players) to receive and interact with web content. UAAG provides guidance on how to make such user agents accessible to people with disabilities (Jacobs, Gunderson, & Hansen, 2002). The interdependencies both between web developers and authoring tools, and between users and user agents, represent what the W3C refers to as the "Essential Components of Web Accessibility" (Henry, 2005b). Together, ATAG, UAAG and WCAG form a trio of standards documents that address the different components of accessible web development.

Of the three web accessibility standards documents published by the W3C, only WCAG 1.0 gained much traction over the subsequent years. Manufacturers were slow to implement the guidance in ATAG 1.0 and UAAG 1.0. This resulted in authoring tools providing very little accessibility support to web developers, and user agents implementing accessibility features either inconsistently or not at all. Referring to this as a “chicken and egg problem”, the W3C highlighted how user agents were less likely to implement accessibility features that most websites did not use (Henry, 2005b). In turn, authoring tools were less likely to implement accessibility features that web developers were not demanding. Finally, web developers were less likely to implement accessibility features that were neither supported by common user agents nor easy to achieve with their authoring tools. According to Meiselwitz, Went, and Lazar (2010), the relatively small number of authoring tool and user agent manufacturers, compared to the vast number of individuals and organisations developing web content, made the dominance of WCAG 1.0 inevitable.

In an effort to encourage the broader adoption of web standards, a “grassroots coalition” of web professionals, known as the Web Standards Project (WaSP), formed an Accessibility Task Force (ATF) in June 2005 (Web Standards Project, n.d.) (not to be confused with the W3C’s HTML Accessibility Task Force (Quin, 2013) which provides a forum for the discussion of accessibility issues in HTML). The WaSP had originally formed in 1998 in response to the so-called “browser wars” of the late 1990s, where rival user agent manufacturers competed to gain market share by implementing proprietary features instead of adhering to web standards. In 2001, with the browser wars largely over, the group turned its attention to, amongst other issues, web accessibility. The aims of the ATF were to work closely with authoring tool and user agent manufacturers to promote web accessibility, to educate and inform web developers, and to establish reliable patterns and methodologies for developing standards-compliant, accessible websites. The efforts of the WaSP culminated in the creation of WaSP InterAct (n.d.), a web curriculum framework based on web standards and best practices. The WaSP continued until 2013 when the group was disbanded, claiming to have achieved its objectives.

Despite WCAG 1.0 being a voluntary set of guidelines developed by a voluntary international body with no regulatory power, it had rapidly become the de facto standard for web accessibility. Due to WCAG 1.0’s close adherence to the technologies of the time, however, coupled with the W3C’s early reluctance to embrace proprietary technologies, the guidelines soon became out-dated. Work began on a second version of the guidelines and, in December 2008, almost a decade after the first version, and following a protracted consultation process, WCAG 2.0 was published as a W3C

Recommendation (Caldwell et al., 2008a). The revised guidelines were organised according to four principles of accessibility that form the acronym POUR:

- Perceivable (users must be able to perceive the information being presented);
- Operable (users must be able to operate the interface);
- Understandable (users must be able to understand both the information being presented as well as how to operate the interface); and
- Robust (users must be able to access the content as technologies advance).

Guidelines under each principle (12 in total) provided the basic goals that web developers should work toward to make content more accessible. Each guideline included a number of success criteria (61 in total). Similar to the concept of checkpoints in WCAG 1.0, success criteria were testable statements that web developers could use to validate conformance. Scrapping the concept of Priority Levels in WCAG 1.0, the WCAG Working Group directly assigned each Success criterion a conformance level: A, AA or AAA. There were 25 Level A, 13 Level AA, and 23 Level AAA success criteria. WCAG 2.0 was published alongside two supporting documents: Techniques for WCAG 2.0 (Caldwell, Cooper, Reid, & Vanderheiden, 2008b) and, perhaps reflecting the increased complexity of the guidelines, Understanding WCAG 2.0 (Caldwell, Cooper, Reid, & Vanderheiden, 2008c).

WCAG 2.0, like its predecessor, had a global influence on accessibility legislation and, in October 2012, was ratified as an official ISO International Standard (ISO/IEC, 2012). At the time of writing, Australia, Canada, the EU, France, Germany, Hong Kong, Iceland, India, Italy, Ireland, Israel, Japan, Netherlands, New Zealand, Norway, and Spain have either adopted WCAG 2.0 in its entirety or created legislation and local standards that incorporate by reference the guidelines (Feingold, 2017; Rogers, 2017). Though the UK's equality legislation (Equality Act 2010) does not specify any technical standards of web accessibility, the UK Government Digital Service (GDS) mandates baseline conformance to WCAG 2.0 Level AA for all government websites (Government Digital Service, 2017). Similarly, the US government encourages organisations with public-facing websites to conform to WCAG 2.0 Level AA to ensure compliance with the Americans with Disabilities Act (ADA) of 2010. A long-awaited "refresh" of the US Section 508 guidelines, approved in September 2016 (United States

Access Board, 2016), will incorporate by reference WCAG 2.0 Level AA ("WCAG 2.0 Conformance", n.d.).

Technical advances in the early 2000s brought an increasing complexity and sophistication to the web (Gibson, 2007). So-called Web 2.0 technologies, such as AJAX (asynchronous JavaScript and XML), allowed websites to behave more like desktop applications, supporting richer, more interactive user experiences (Garrett, 2005). Unlike Web 1.0 websites, which were characterised by collections of static web documents, Web 2.0 websites provided opportunities for users to interact and create content themselves. Web content could be updated dynamically without the need to reload entire webpages. Web developers could create custom user interface controls beyond those available in regular HTML, such as sliders, dials, and colour pickers, and incorporate advanced interaction modes, such as drag-and-drop functionality. However, ATs, such as screen readers, had been designed around the assumption that, once loaded, webpages would remain static and unchanging. Consequently, they were unable to process the dynamically generated content of Web 2.0 websites and struggled to interpret custom user interface controls and advanced interaction modes. The rush to embrace new Web technologies without ensuring their compatibility with ATs had resulted in a substantial "accessibility gap" for people with disabilities (King, Nurthern, Bijl, & Cooper, 2017).

To accommodate the shortfall in accessibility caused by the emergence of Web 2.0 technologies, the WAI developed a raft of new guidelines and specifications. For instance, WAI-ARIA (Web Accessibility Initiative-Accessible Rich Internet Applications), which was published as a W3C recommendation in March 2014 (Craig & Cooper, 2014), extended existing HTML to provide the additional semantic information necessary for ATs to handle dynamic content and custom user interface controls. Many of the concepts introduced in WAI-ARIA were incorporated into the fifth and current version of the HTML specification, which was published as a W3C Recommendation in October 2014 (Hickson et al., 2014). HTML5 incorporated a range of new semantic and structural elements for marking up web content in a more accessible way. Authoring tool and user agent support for these new technologies was also stimulated by the development of ATAG 2.0, which the W3C published as a Recommendation in September 2015 (Richards, Spellman, & Treviranus, 2015), and UAAG 2.0, which the W3C published as a Working Group Note in December 2015 (Allan, Lowney, Patch, &

Spellman, 2015). At the time of writing, the WAI is preparing to release a minor update to WCAG 2.0, and has begun discussions about a major revision of the guidelines (known informally as Silver), which may incorporate guidance from both ATAG and UAAG (W3C/WCAG working Group, 2016). Following a recent recharter, the WAI EOWG has also turned its attention towards developing materials for web developers to improve their understanding and implementation of web accessibility (Henry, 2017a). These materials include *Easy Checks*, which offer guidance on how to evaluate web accessibility, *Web Accessibility Perspectives*, which convey the benefits of web accessibility through brief engaging text and video stories, and *Web Accessibility Tutorials*, which describe how to create WCAG-conformant websites (Henry, 2017b). Web developers also continue to be supported by a range of web accessibility evaluation tools that address a range of requirements in a range of environments (Eggert & Abou-Zahra, 2016).

The numerous web accessibility initiatives that have emerged over approximately the last two decades have mirrored and reacted to the rapid growth and development of the web. Web developers now have access to a suite of technologies for creating sophisticated websites and are supported by a variety of tools and a well-established body of accessibility information. The definition of web accessibility standards, policies and legal imperatives further serves to support, encourage and compel web developers to create accessible websites.

2.3. The impact of web accessibility initiatives

In parallel to various web accessibility initiatives over approximately the last two decades, there have also been numerous attempts to take a measure of the current state of web accessibility. Such studies typically pick a selection of websites (often the most popular at the time) but sometimes they might be from a particular sector, such as the government. They then evaluate the websites, using a variety of methods (including automatic, manual, and, sometimes, user testing: see Section 2.1) and following a variety of guidelines (see Section 2.2), to present a measure or benchmark of web accessibility. This has resulted in a considerable amount of literature that collectively measures the accessibility of websites over time.

In Chapter 3, I present a systematic review of 397 web accessibility evaluation studies published over a 15-year period between 1999 and 2014. The aims of this review are to investigate how empirical research on web accessibility is conducted, understand the changing state of the art in the field, and determine whether the accessibility of websites has improved over time. The study shows a persistent occurrence of accessibility problems that, despite the introduction of various web accessibility initiatives and irrespective of the refinement of accessibility guidelines, does not appear to be improving.

2.4. Attitudes and awareness towards web accessibility

The best efforts and good intentions of researchers, practitioners and lawmakers over the last twenty years have culminated in a variety of web accessibility tools, guidelines and resources. Given this abundance of support material, it remains unclear why web developers still struggle to create accessible websites. In an attempt to explain the discrepancy, numerous authors have surveyed attitudes and awareness towards web accessibility among web developers.

In one of the earliest attitudinal studies of web developers, Vora (1998) surveyed 138 web authors and designers to understand how they approach their work. Entitled "Designing for the Web", the survey's ultimate aim was to inform the creation of usable web development tools. It covered a range of topics, including professional background and experience; the size and type of websites developed; the development tools used; the design style guides used; and methods of publishing webpages. While the survey provided an effective "snapshot" of web development in the late 1990s, it only briefly broached the topic of web accessibility (although considering that Vora conducted the survey at a time when web accessibility was only just gaining attention, such an omission is perhaps understandable). With WCAG 1.0 still a year from publication, Vora directed interested readers to the work emerging from the W3C's nascent WAI. Nevertheless, the survey's findings were clear: while only 16.7% of respondents claimed to have "always" designed webpages for accessibility, 50% claimed to have "sometimes" optimised their designs to improve web accessibility, and 33.3% claimed to have "never" considered web accessibility at all.

Six years later, Lazar et al. (2004) conducted a survey of 175 webmasters, to explore their knowledge and perceptions of web accessibility. The survey comprised fifteen multiple-choice and open-ended questions, exploring familiarity with web accessibility laws and guidelines; awareness of accessibility tools and ATs; personal and organisational attitudes to web accessibility; and perceptions of when and why websites should or should not be accessible. Despite many of the questions focusing specifically on accessibility for users with visual impairments, the survey provided a broad understanding of attitudes and awareness towards web accessibility in general. The results, at first, indicated a reasonable familiarity with web accessibility and suggested that numerous initiatives to encourage and promote web accessibility had been successful. For example, 73.7% of respondents claimed to be familiar with the US Section 508 guidelines, and 64% claimed to be familiar with WCAG 1.0 (both of which had been available for several years at the time the survey was conducted). Furthermore, 78.9% of respondents claimed to be familiar with accessibility evaluation tools and 69.1% even claimed to have used such tools. When asked about whether they implemented web accessibility, however, only 56% of respondents indicated that their current website was accessible and only 58.8% claimed that their organisation had plans to make their websites accessible in the future. These findings seem at odds with the respondents' reported familiarity with web accessibility guidelines and tools and suggest that they were failing to translate their awareness of web accessibility into implementation. It also hints at the possibility of social desirability bias (Nederhof, 1985), with respondents providing what they believe to be desirable but inaccurate answers in an effort to be viewed favourably by others.

At a similar time to the survey by Lazar et al., the UK Disability Rights Commission (DRC) commissioned a survey of web developers to assess their knowledge of and interest in web accessibility (Disability Rights Commission, 2004). Unfortunately, despite the authors approaching 388 web development agencies to complete a short postal questionnaire, the majority failed to respond. The 23 agencies that did respond, however, reported reasonably positive attitudes to web accessibility. While only 9% of respondents claimed any sort of web accessibility expertise, 80% reported that their agency attempted to develop accessible websites at least some of the time. Almost two thirds of respondents (65%) were familiar with WCAG 1.0 and generally found the guidelines useful, though, surprisingly, far fewer (21%) were familiar with automated accessibility evaluation tools. While 70% of respondents claimed to have conducted user testing, only 9% had ever included disabled users in such tests. Follow-up face-to-face interviews with 25 web developers largely complemented these findings: 81% of web developers claimed to be aware of web accessibility and 58% claimed they would always broach the subject with clients (though in only 31% of cases did they receive a positive response). Web developers' familiarity with accessibility guidelines and testing was slightly lower in the interviews than in the questionnaire responses, with only 48% claiming to have ever used guidelines or conducted automated testing, and only one web developer claiming to have ever involved disabled people in user testing.

One year later, Rosson, Ballin, Rode and Toward (2005) revisited the "Designing for the Web" survey first conducted by Vora (1998). Broadening the scope of recruitment to include both professionals and amateurs alike, Rosson et al. surveyed 334 web developers to understand the needs, problems and processes that they followed. The online survey, which the authors based loosely on Vora's original survey, comprised 37 questions on a range of topics, including: web development experiences and difficulties encountered; web development tools and testing; and web development working styles. As with Vora's (1998) earlier work, the survey provided an effective "snapshot" of web development in the early 2000s. It is disappointing, however, that web accessibility was still only briefly touched upon in the survey. Rosson et al. made no attempt to gauge web developers' awareness of web accessibility but, instead, asked them how often (on a scale of 1 to 5, with 1 being never, and 5 being always) they tested for web accessibility. The respondents' mean rating to this question was 2.75 (standard deviation (SD): 1.41, number (N): 276), suggesting that they tested fairly infrequently for web accessibility, if at all. The change in how the authors measured this question (from a simple yes/no agreement to a 5-item rating scale) prevents direct comparison with Vora (1998). However, seven years later, the findings were disappointingly consistent: web developers were still neglecting to make their websites accessible.

The following year, the EU-funded BenToWeb (Benchmarking Tools and Methods for the Web) project, (Petrie, 2006) conducted surveys and interviews of 249 European web developers to gain an understanding of attitudes and awareness toward web accessibility. The surveys and interviews covered a range of topics, including: development languages, tools and environments used; organisational attitudes to web accessibility; and web accessibility knowledge and practice. The survey findings suggested both a growing awareness and greater implementation of web accessibility among web developers. Almost 70% of the respondents (69.1%) reported that their organisation was either "interested" or "very interested" in web accessibility. Respondents were also reasonably confident about their understanding of web accessibility, with nearly 39.8% claiming to be "knowledgeable". Furthermore, 73.9% of respondents claimed to have tried to develop an accessible website and 67.4% of respondents claimed to have conducted some form of accessibility testing. Curiously, only 39.8% claimed to use WCAG, which seems at odds with the proportion of respondents that claimed to have conducted accessibility testing and suggests that some were undertaking testing without the support of accessibility guidelines. Nevertheless, the findings indicated that web developers were beginning to put into practice their newfound awareness of web accessibility.

Two years later, Freire et al. (2008) conducted an exploratory survey of 613 Brazilian web development professionals to investigate their awareness and understanding of web accessibility. The authors recruited respondents from a broad range of academic, industrial and governmental roles in Brazil. The survey comprised seventeen closed and open-ended questions covering a range of topics, including: knowledge of HTML and CSS; awareness of disabled people and how they use the web; knowledge of ATs; knowledge of Brazilian accessibility legislation; knowledge and use of accessibility evaluation techniques; attitudes towards web accessibility; and reasons for considering or not web accessibility in web development. The survey also asked respondents to outline their opinion on the importance of web accessibility and to provide suggestions on how to improve web accessibility in Brazil. Unfortunately, the results indicated only a mild awareness of web accessibility among web development professionals in this part of the world. For example, 45.2% of respondents stated that they were aware of technologies to make the web accessible to blind users, but that they did not know how to actually create accessible webpages. Just over a quarter of respondents (27.6%) claimed to have heard about the possibility of blind people using the web but were not sure how it was possible. Almost two fifths (39.2%) of respondents had no awareness of WCAG and 30.2% only had a basic knowledge it. Similar percentages (40.3% and 33% respectively) were recorded for awareness of a Brazilian-specific accessibility law. This relatively mild awareness of accessibility issues and guidelines was also reflected in the respondents' reported web development practices. Just under half of the respondents (47.8%) claimed not to use any accessibility evaluation methods. Only 19.9% of respondents stated that they consider web accessibility, 44.4% considered it partially, and 35.4% did not consider web accessibility at all.

The following year, the EU-funded ACCESSIBLE (Accessibility Assessment Simulation Environment for New Applications Design and Development) project

(Korn, 2009) conducted a survey of 254 European web developers to determine why they often fail to incorporate web accessibility and how best to support them. The survey comprised both closed- and open-ended questions covering a range of topics, including: web development background; comprehension of web accessibility; familiarity with web accessibility; and expectations of web accessibility. It is important to note that only 65% were actual web developers and that 25% were company directors, and 10% were assistants or students. The results indicated a much greater awareness of web accessibility among this cohort of respondents, with over 90% claiming to be “acquainted” with accessibility (i.e. they were familiar with the term but did not necessarily know its exact meaning). Respondents’ awareness of web accessibility standards and guidelines was mixed: although 87% claimed to be familiar with WCAG 1.0, only 45% claimed to be familiar with WCAG 2.0 (which was published two years previously). Even fewer respondents were familiar with Section 508 (39%) and ATAG 1.0 (7%), although 65% were familiar with the WAI-ARIA specification (which, at the time, was only available in draft form). The project team suggested that respondents’ relatively low familiarity with certain standards was due to the broad range of standards and guidelines available. The results of the ACCESSIBLE survey were fairly encouraging and appeared to continue the trend of growing awareness of web accessibility observed in previous surveys. The flexibility allowed in some of the questions, however, may have resulted in a more generous assessment of the respondents’ awareness. For example, a question that explored respondents’ familiarity with accessibility was suffixed by the disclaimer: “It’s not necessary to know the exact meaning of the term”. This could have allowed for a very loose awareness of accessibility and may reflect why over 90% responded positively. Similarly, as the authors acknowledged, “being aware of standards, does not mean [respondents] actually know the standards and can implement this” (p. 43). Unfortunately, the authors failed to explore this distinction any further or investigate whether web developers had translated their apparent awareness of web accessibility into implementation.

A year later, Trewin, Cragun, Swart, Brezin and Richards (2010) conducted a survey of 48 software engineers/developers (and 1 information architect) working for IBM, to investigate their attitudes and awareness towards web accessibility. The survey comprised 23 closed- and open-ended questions on a range of topics, including level of expertise in accessibility; familiarity with web accessibility resources and accessibility evaluation tools; difficulties faced and the amount of time taken to produce accessible products; barriers to accessibility; and the relative importance of accessibility tool features. The respondents' familiarity with accessibility was fairly evenly distributed: fourteen claimed to be accessibility experts, fifteen claimed to have an intermediate knowledge of accessibility, and a further fourteen claimed to be accessibility novices. It is important to note, however, that Trewin et al. specifically recruited web developers with at least some awareness of web accessibility. Three quarters (75%) of respondents claimed to have developed or tested a product for accessibility. They also admitted being heavily reliant upon accessibility evaluation tools: 69% of respondents reportedly used both an automated evaluation tool and an AT; 16% used only an AT; 14% used only an automated evaluation tool; and 8% did not use any accessibility evaluation tool. The majority of respondents (68%) relied on internal IBM checklists and guidelines to understand and resolve accessibility issues. Just over a third (35%) referred to W3C documentation and 14% turned to Google to locate accessibility information. Other resources they reportedly used included the Section 508 documentation as well as technology-specific guidelines. An obvious limitation of the survey is that it was completely IBM-oriented, focusing on IBM developers who work to IBM guidelines following IBM corporate instructions. The motivations and experiences of this cohort may be very different from those not working for one of the world's largest IT companies. Nevertheless, the results summarise the difficulties in implementing web accessibility in a heavily corporate environment.

Two years later, Putnam et al. (2012) conducted an online survey of 185 UX and Human-Computer Interaction (HCI) professionals (including web developers) from the US, UK, Brazil, Germany and China, to explore their attitudes and approach to accessibility. The survey was part of a wider body of research exploring how UX and HCI professionals consider accessibility in their work. The short survey required respondents to answer two questions relating to accessibility: the first required respondents to rate how important they felt it was to make technology accessible; the second, open-ended question asked: "How do you consider accessibility in your work? In other words, what types of efforts/research is performed to help make the products/services you are involved in creating accessible to diverse users including people with disabilities?" The survey results indicated a positive approach to web accessibility, at least among certain groups of professionals. The majority of respondents (70%) claimed to give some consideration to accessibility. This group of respondents also tended to rate accessibility as "important" or "very important" in their work. Reported efforts by this group included making "special accommodations" for people with disabilities (such as providing alternative text on images, increasing the size of interface elements, or including subtitles on videos); conducting some form of accessibility testing; consulting with accessibility experts; referring to laws, guidelines and best practices; being an accessibility advocate within their organisation; and making other accessibility considerations (such as low bandwidth, slow CPU and smaller screen sizes). The other 30% of respondents (which, unfortunately, included a large number of web developers) gave no consideration to accessibility whatsoever and rated it as much less important in their work. A subset of this group (19%), however, expressed regret or was apologetic about not considering accessibility, and a further subset (9%) conveyed signs of hope and progress towards accessibility.

Three years later, in an attempt to harmonise understanding of web accessibility, Yesilada et al. (2015), surveyed 300 people with "an interest in accessibility" about their perceptions of the topic. The authors recruited a mixture of technical and non-technical stakeholders (including web developers) from a range of academic, industrial and governmental roles in many countries around the world. Respondents ranked their agreement with 33 statements that explored the relationship between accessibility, user experience, and usability; inclusion and exclusion; and web accessibility evaluation. The survey results indicated a more refined, nuanced understanding of web accessibility. Respondents considered web accessibility to be strongly related to usability and necessary for a good user experience. They also understood that web accessibility benefits all types of people and not just people with disabilities. Beliefs about the main drivers of web accessibility varied: few believed that web accessibility was driven by business revenue, though respondents were divided on the motivational role of regulatory legislation. The majority of respondents believed that web accessibility should be achieved not by rapid testing but by a user-centred design process. While many found WCAG helpful, they considered training necessary to use the guidelines effectively. Respondents believed that practitioners should not solely rely on guidelines to determine the accessibility of websites and that more valid and reliable results could be achieved through user testing. Yesilada et al. (2015) drew upon these results to argue for a much broader definition of accessibility that encompasses not only users' abilities but also the technology they use, the situation they are in, and their context of use. To this end, they proposed several practical implications, including better training and education; greater communication of accessibility; and further empirical research to explore unresolved issues that emerged from the survey, such as the consequences of using different evaluation techniques in assessing web accessibility. Although the authors made some attempt in their analysis to distinguish between technical and non-technical respondents, the broad spectrum of respondents makes it difficult to specifically tease out the attitudes and awareness of web developers. Consequently, it is unclear whether, and to what extent, web developers' views towards web accessibility had changed. Furthermore, the use of (by the authors' admission) occasionally ambiguous agreement statements to solicit respondents' views, without any opportunity for them to provide additional comments, precluded a more comprehensive interpretation of the findings. Nevertheless, the survey provided a further "snapshot' of web development in 2015, indicating a mature attitude and greater awareness towards web accessibility.

One of the apparent trends observed in the surveys of web developers over the last twenty years is a growing awareness, both of web accessibility and of associated guidelines, tools and legislation. Web developers now claim to be more familiar with WCAG, Section 508, and the needs of disabled people in general than they were at the turn of the century. Some web developers also express regret or embarrassment when, for whatever reason, web accessibility cannot be achieved. Although this indicates some degree of social responsibility among web developers, it is important to acknowledge that respondents to such surveys decide entirely for themselves whether or not to respond, and therefore some degree of self-selection bias is inevitable. Web developers who are more interested in and more engaged with web accessibility are inevitably more likely to respond to surveys that solicit their views on the topic. Furthermore, the anonymous environment of (often online) surveys enables respondents to provide more socially desirable responses and to portray themselves in a more positive light. Nevertheless, given the considerable efforts over the last twenty years by organisations such as the WAI to raise the profile of web accessibility, it is not surprising to witness a growing awareness of the topic among web developers. A more concerning trend to emerge from the surveys, however, is that, despite their greater awareness of web accessibility, web developers appear no more likely to create accessible websites. This trend is corroborated by the considerable number of studies over the last twenty years that indicate little improvement in the accessibility of websites (see Chapter 3).

2.5. Influences on and barriers to web accessibility

Acknowledging that web accessibility can be incorporated (or omitted) at various points in the web development process, Lazar et al. (2004) proposed a Web Accessibility Integration Model. The aim of the model was to draw attention to the various aspects of web development that contribute to the accessibility (or inaccessibility) of websites (see Figure 2.6). The model includes three broad categories of influences on web accessibility: *societal foundations*, *stakeholder perceptions*, and *web development*.

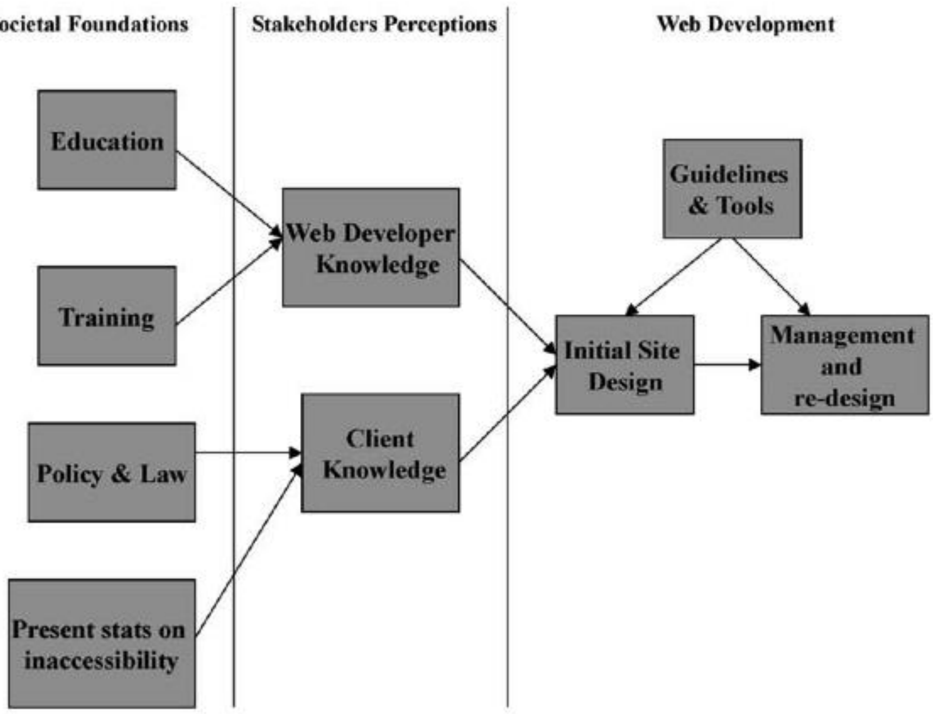


Figure 2.6: The Web Accessibility Integration Model. Reprinted from Fig. 1 in Lazar, J., Dudley-Sponaugle, A., & Greenidge, K. (2004). Web accessibility: a study of webmaster perceptions. *Computers in Human Behaviour*, 20(2), 269-288 with permission from Elsevier.

The first category, *societal foundations,* refers to the value that society places on web accessibility, which, according to Lazar et al., is reflected in both education and training in web accessibility, policy and law relating to web accessibility, and, debatably, the impact of "shocking" accessibility statistics on raising awareness of web accessibility. While these factors undoubtedly contribute to societal awareness of web accessibility, they present a rather simplistic perspective of the subject, one in which web accessibility is limited only by a lack of knowledge and motivated only by guilt and fear of legal consequences. For example, despite the issue emerging from their accompanying survey of webmasters, Lazar et al. neglect to include the impact of market forces on web accessibility. General misconceptions about the needs of disabled people, fuelled perhaps by a lack of advocacy for web accessibility, may contribute to a societal lack of awareness. This, coupled with the limited involvement of disabled people in user testing (revealed in several of the studies described in Section 2.4) and reluctance on the part of disabled people (or a lack of knowledge on how) to complain about inaccessible websites, may lead to a *perceived* lack of demand and, ultimately, a limited budget for web accessibility.

Although Lazar et al. entitle the second category *stakeholder perceptions*, they include only a limited subset of stakeholders in their model, specifically web developers and clients. It is these stakeholders who, they claim, have the greatest influence on whether or not a website will be made accessible. While to some extent this may be true, it neglects the involvement of numerous other stakeholders in the value chain of accessible web development (see Section 1.1), Also, despite referring to stakeholder "perceptions" in the category title, Lazar et al. confusingly refer to web developer and client "knowledge" in the diagram. One's perception of web accessibility—however well informed—does not necessarily correspond to one's knowledge of the subject (a point I demonstrate in contextual inquiries with web developers in Chapter 4). From the text accompanying the diagram, it is clearer that Lazar et al. are referring to the attitudes and perceptions of web developers and clients, which are informed by the different factors in the previous category.

The final category, *web development,* highlights the influence of accessibility guidelines and tools, which affect both the initial website design as well as on-going maintenance and redesign. While the significance of accessibility guidelines and tools may be true, the causality proposed by the model is questionable. Lazar et al. suggest that both societal foundations and stakeholder perceptions influence actual web development. However, they then go on to suggest that the guidelines and tools used in web development somehow inform the current "working definition" of web accessibility, that is, a tentative, practical understanding that emerges in lieu of a more authoritative definition (which, surely, the guidelines and tools provide?) Putting these concerns aside, the category appears to be more a reflection on the usability of guidelines and tools, with Lazar et al. suggesting that the easier they are to use and understand, the more likely web developers are to incorporate accessibility into their websites.

Although the Web Accessibility Integration Model by Lazar et al. (2004) draws attention to various aspects of web development that contribute to the accessibility (or inaccessibility) of websites it is limited in several ways. Recognising some of these limitations, and following interviews with 23 website practitioners, Farrelly (2011) proposed an Expanded Web Accessibility Integration Model, which attempts to provide a more "comprehensive and holistic" representation of influences on web accessibility (see Figure 2.7). It incorporates a number of "pivotal" factors absent from the original model by Lazar et al., such as attitudes towards disability, market forces, and customer demand and advocacy. It also expands the web development category to include guidelines, support materials, authoring tools, testing support and hired experts. Farrelly also adds a further category, entitled *end user*, comprising transcoding systems, user agents, and ATs. This category appears to place some of the responsibility for web accessibility on end-user technologies but, regrettably, the accompany text does not elaborate on this.

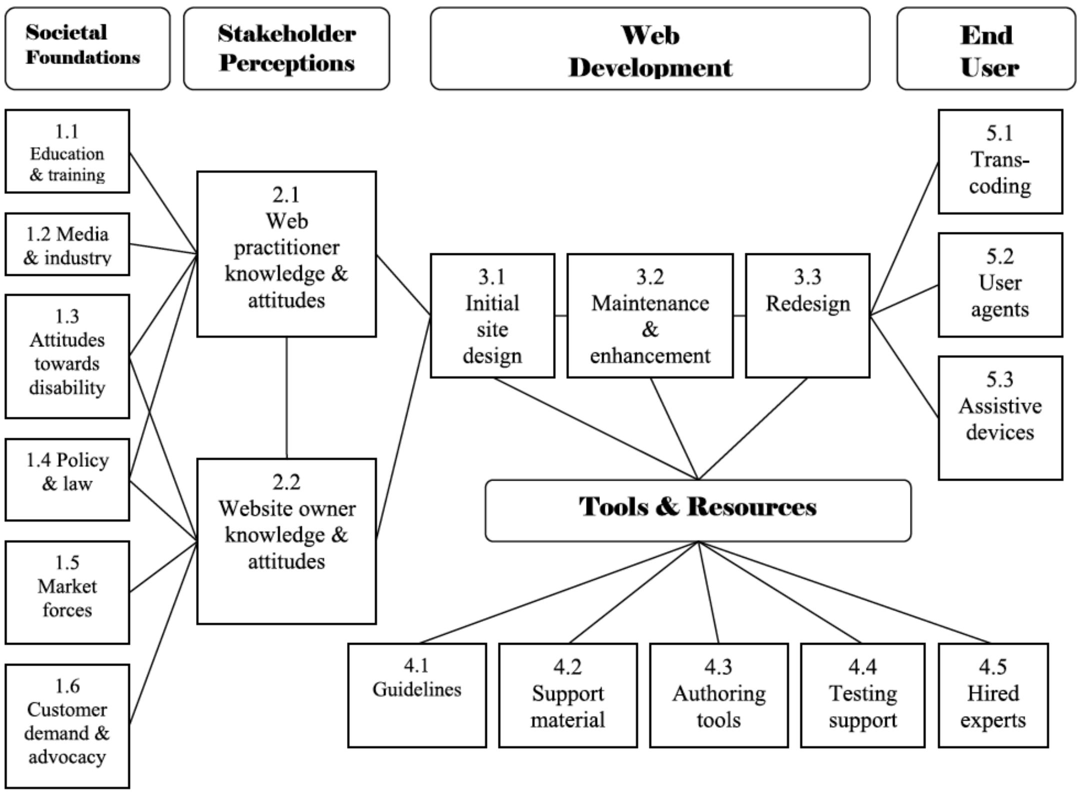


Figure 2.7: The Expanded Web Accessibility Integration Model. Reprinted from Fig. 1 in Farrelly, G. (2011). Practitioner barriers to diffusion and implementation of web accessibility. *Technology and Disability*, 23 (4), 223-232 with permission from IOS Press.

While the Expanded Web Accessibility Integration Model provides a more complete perspective of web accessibility, it suffers from some of the limitation of its predecessor. Its focus is still predominantly on a limited subset of stakeholders (this time, "web practitioners" and "website owners"); perceptions and knowledge are, again, treated synonymously; and the causality and relative importance of the different categories is even less clear. Nevertheless, the two models provide a useful context for assessing the literature in this field and bringing greater clarity to the apparent disparity between web accessibility awareness and implementation.

2.5.1. Societal foundations

In addition to surveying attitudes and awareness towards web accessibility among web developers, many of the surveys discussed in Section 2.4 examined perceived barriers to accessible websites. Few respondents to these surveys disputed the influence of education and training on web accessibility with many highlighting the need for a thorough grounding in the subject.

Webmasters surveyed by Lazar et al. (2004) cited the technical challenge of creating accessible websites, coupled with insufficient training and education, as major barriers to web accessibility. Lazar et al. observed that web accessibility did not feature in any national curriculum in computer science or related discipline and that few organisations outside of government offered web accessibility training. The report commissioned by the DRC (2004) also called for modules on disability awareness and accessibility to form part of basic web development training. The majority (85%) of web developers surveyed as part of the ACCESSIBLE project (Korn, 2009) also indicated a need for advanced education in the area of accessibility.

In surveys and interviews with 117 designers, information officers and accessibility advocates by Knight (2003), respondents considered the biggest barrier to web accessibility to be a lack of education, with 86% agreeing that "developers do not have adequate training". Web developers interviewed as part of the BenToWeb project (Petrie, 2006) expressed a lack of expertise or knowledge about web accessibility, which they also attributed to insufficient education and training. A lack of training was also one of the reasons given by the Brazilian web professionals surveyed by Freire et al. (2008) as to why they do not include web accessibility in their projects. They called for more intensive promotion of web accessibility and suggested the topic should feature more prominently in organisational training and web-related educational courses.

Web practitioners interviewed by Farrelly (2011) drew attention to the recruitment problems posed by the lack of education and training and highlighted how organisations often struggle to locate sufficiently skilled web developers. A Delphi study conducted by Hong, Trimi, Kim and Hyun (2015) with 30 web accessibility experts (including some web developers) identified a lack of accessibility education as one of the main factors hindering web accessibility. Finally, improved training and education in web accessibility was also one of the recommendations made by Yesilada et al. (2015) in response to their survey of 300 web professionals.

Many authors have explored how education and training for web developers might incorporate web accessibility. Ludi (2002) outlines the pedagogical and practical considerations of incorporating web accessibility into computer science curricula, with the aim of ensuring prospective web developers are adequately equipped with the appropriate knowledge, skills and attitudes. Rosmaita (2006) proposes an "accessibility first" approach to web design education, arguing that web accessibility should be brought "front and center and made the focus of the course", instead of being relegated to a web development "add-on".

Recognising that web accessibility is rarely taught to any depth, Ortner, Batušić, and Miesenberger (2004) report the development a postgraduate e-learning course at the University of Linz in Austria, called *Barrierefreies Webdesign* (Accessible Web Design). The two-year course, which was aimed predominately at web designers, follows a Design For All approach and includes topics such as: technical fundamentals; ATs; guidelines and laws; accessibility; design and usability; and practical experience.

Gellenbeck (2005) outlines a strategy developed at the Central Washington University in the US for integrating web accessibility into computer science curricula. The strategy emphasises the early introduction and continued discussion of accessibility issues. Although Gellenbeck does not report on the effectiveness of the strategy, he notes that it was well received by students and commercial partners alike.

Harrison (2005) reports on the introduction of web accessibility, and, specifically, designing for screen reader users, into an undergraduate web design course at the

University of Wisconsin-Eau Claire in the US. Harrison successfully evaluated the effectiveness of this intervention by instructing students to create an accessible website using the skills they had gained.

Benavídez, Fuertes, Gutiérrez, and Martínez (2006) discuss the use of two key materials in teaching web accessibility at the Technical University of Madrid in Spain: an automatic accessibility evaluation tool, called *HERA*, and a fictitious and highly inaccessible website, called *Contramano*. Similarly, Freire, Paiva, Turine and Fortes (2007) describe the use of screen readers in a one-day introductory course on web accessibility at a public university in Brazil to make students aware of the problems that disabled users face in using the web.

Efforts to incorporate web accessibility into education and training are laudable and would undoubtedly have a bigger influence on the accessibility of websites, were it not for the fact that many web developers are self-taught and learn on the job.

Vora's survey (1998) found that web developers come from a variety of academic and professional backgrounds, including "computer programming, graphic design, human factors marketing/advertising, user interface design, and technical writing, and often have additional training in cognitive psychology multimedia, and education" (p. 3). Professional and amateur web developers surveyed by Rosson et al. (2005), who were often self-taught, claimed they would gain new web development skills by consulting "FAQs, books, and tutorials" and "examples of similar sites from which you can get ideas and copy code" and "a friend or coworker who knows how to do it" but rarely from more formal educational sources, such as software help materials, seminars, or support hotlines.

More than half of respondents to the survey by Freire et al. (2008) had never had any formal web accessibility training. Those that did have some awareness of web accessibility had largely gained it either on the job or through self-learning. Despite calling for formal web accessibility education to be enforced in computer science programs, Hong et al. (2015) acknowledge the majority of web developers are completely self-taught.

Although many of the website practitioners interviewed by Farrelly (2011) had received formal training in their profession, they reported that web accessibility was largely ignored or insufficiently addressed. Similarly, of the 42.2% of web developers surveyed by the BenToWeb project (Petrie, 2006) who had received any formal training in web design, nearly half of these (41%) said there had been no material on web accessibility. In follow-up interviews, several participants highlighted how web developers have varying backgrounds in design, graphic design, and programming and that it is difficult to reach all of these groups.

Though authors (and web developers) are fairly unanimous about the role of education and training in web accessibility, they are more divided about the influence of policy and law. The enshrinement of web accessibility in policy and legislation (such as Section 508 in the US, outlined in Section 2.1) is undoubtedly a powerful motivator for web developers and their clients to ensure their websites are accessible. Without proper enforcement, however, the effectiveness of web accessibility policy and law is limited.

Respondents to the survey of 175 webmasters by Lazar et al. (2004) indicated that stricter government regulations would have the biggest influence on the accessibility of their websites. Respondents and participants in Knight's (2003) surveys and interviews with 117 web designers, information officers and accessibility advocates ranked "a lack of legal action" as the fourth biggest barrier to developing accessible websites.

A study of the accessibility of 100 US federal government homepages, by Olalere and Lazar (2011) infers that government agencies are unlikely to make their websites Section 508-compliant without government oversight, data collection, and enforcement. The Delphi study conducted by Hong et al. (2015) identified a lack of statutory or financial support from the government as one of the main factors hindering web accessibility. They highlight how organisations lack the willingness to comply when existing laws regarding web accessibility are not compulsory.

For web accessibility policy and law to be effective, web developers and clients must also know it exists and be aware of its significance. The survey of 613 Brazilian web development professionals by Freire et al. (2008) highlights how few respondents were aware of web accessibility legislation and even fewer actually adhered to its requirements.

Effective web accessibility policy and law must also strike a balance between "carrot" and "stick". Although some web developers consider the lack of punishment a barrier to web accessibility, others find the reward for accessible web development insufficient. To this effect, respondents to the survey by Yesilada et al. (2015) were divided on the motivational role of regulatory legislation. Similarly, several web developers interviewed as part of the BenToWeb project (Petrie, 2006) considered web accessibility legislation to be of limited value. They argued that legal enforcement motivates web developers to meet only the minimum requirements for accessibility conformance and saps any incentive to do more.

The influence of "shocking" statistics on web accessibility is difficult to ascertain. Certainly, a greater awareness of the difficulties that disabled people face in accessing websites would motivate many web developers. Web developers surveyed by Lazar et al. (2004) claimed they would be more inclined to make their websites accessible if they knew that people with disabilities were actually using them. Lazar et al. argue that mainstream media reports about the inaccessibility of websites contribute to stakeholders' perceptions of web accessibility and serve to shame them into taking action. They lament, however, that such statistics "do not seem to influence people to make more web sites accessible or change the patterns of education" (p. 3).

Although Lazar et al. find this contradiction surprising, they do not appear to consider the possibility that stakeholders may be simply unaware or unfamiliar with the problems faced by people with disabilities. The web practitioners interviewed by Farrelly (2011) criticised the lack of media coverage of disabled people. Many felt that disability was inadequately covered in popular or trade media other than by presenting stereotypes. Farrelly argues that this lack of coverage contributes to the ignorance of the public and reinforces beliefs that accessibility can be ignored.

A considerable number of authors have attempted to take a measure of the current state of accessibility (see Chapter 3), but rarely are the outcomes of such studies fed back to website stakeholders or reported to a wider audience through mainstream media. A notable exception was a landmark accessibility investigation of 1000 UK homepages by the DRC (2004), which found the majority of websites failed to conform to even the most basic level of accessibility, and which was reported by the UK media (e.g. Adams-Spink, 2004). Similarly, a report by the UK Cabinet Office (UK Cabinet Office, 2005), which revealed that almost every European government website failed to meet basic standards of accessibility, received some media coverage (e.g. "Government sites fail web tests", 2005). Web accessibility is increasingly prevalent in trade media but rarely breaks through to the mass media. The publication of accessibility standards, policies and legislation, such as WCAG 1.0, WCAG 2.0 and the recent British Standard 8878, have received some media coverage (e.g. Rosencrance, 2001; Adams-Spink, 2006, 2008). Though accessibility statistics may indeed be "shocking", their influence on stakeholders' perceptions of web accessibility will be limited without proper communication.

2.5.2. Stakeholder perceptions

Both Lazar et al. (2004) and Farrelly (2011), in their respective models, consider the societal foundations described in the previous section to be influential in shaping the perceptions of stakeholders involved in web development. Lazar et al. (2004) argue that if neither web developers nor clients are aware of or passionate about web accessibility then the resulting website is unlikely to be accessible. Gregor, Sloan and Newell (2005) also consider the behaviour of designers and developers to be heavily motivated by their attitudes towards accessibility. They argue that negative attitudes towards accessibility can result in designers and developers producing content that is inaccessible or insufficiently accessible. Similarly, Rosson et al. (2005) suggest that web accessibility is often overlooked due to a lack of awareness or concern on the part of web developers.

Many of the surveys discussed in Section 2.4 characterised web developers as being responsive to, compassionate towards, and increasingly aware of the needs of people with disabilities. Farrelly (2011) also emphasises how each of the web practitioners he interviewed expressed positive attitudes towards disabled people and web accessibility and "appeared to care earnestly about this issue and wanted to learn more about web accessibility" (p. 6). However, as also discussed, the opportunity for self-selection and social desirability biases that surveys and interviews create, results in them rarely portraying web developers in anything but a positive light. Consequently, it is difficult to ascertain how sincere their responses are and what influence they might have on the accessibility of websites. The experts involved in the Delphi study by Hong et al. (2015), who comprised designers, consultants and educators as well as web developers, provide a perhaps slightly more objective perspective. They characterise web developers as having "passive, complacent attitudes" towards web accessibility and being motivated only by web accessibility compliance.

Web developers' attitudes towards web accessibility appear to be largely influenced by those of the various other stakeholders involved in the value chain of accessible web development (see Section 1.1), from whom, it would appear, they face considerable resistance. Many respondents to the surveys discussed in Section 2.4 highlighted the prohibitive influence of clients, management and other stakeholders, who reportedly have little awareness and no clear understanding of what accessible web development entails. With almost two thirds of respondents (64%) agreeing that "management is unaware of the importance of web accessibility", Knight (2003) considered the second biggest web accessibility barrier to be a lack of awareness among management. Common misconceptions about the costs and challenges involved in making accessible websites lead to web accessibility being portrayed as a confusing, overwhelming and unaffordable 'optional extra'.

In discussing potential barriers to web accessibility, some respondents to the surveys discussed in Section 2.4 alluded to conflicts of attitudes among their immediate colleagues. Respondents to the DRC (2004) survey reported conflicts between accessibility and aesthetic considerations, reinforced by increased client demands for graphics and other creative (but often inaccessible) enhancements. Similarly, respondents to the surveys by Lazar et al. (2004) and the BenToWeb project (Petrie, 2006) reported facing resistance from graphic designers unwilling to compromise their artistic vision in support of web accessibility. The Delphi study conducted by Hong et al. (2015) also identified concerns over simple designs as one of the main factors hindering web accessibility. Complying with accessibility standards was seen as something that would result in plain designs. Conversely, "fancy and dynamic websites" were considered antithetical to web accessibility.

Despite their best intentions, respondents to the survey by Lazar et al. (2004) claimed they often struggle to convince stakeholders of the importance of web accessibility. Similarly, the DRC (2004), the BenToWeb project (Petrie, 2006), and Farrelly (2011) highlight the difficulty that web developers face in persuading clients to take web accessibility seriously. Some web developers attempt to pique clients' attention by emphasising the ancillary benefits of web accessibility, such as improving operability, SEO, and increased potential audience. Rarely are such attempts successful, however, as clients often have little knowledge about, or interest in, the topic (Disability Rights

Commission, 2004). Instead, web developers reluctantly tend to ignore, postpone or comply with a lower level of web accessibility (Farrelly, 2011).

2.5.3. Web development

Although the attitudes and perceptions of stakeholders undoubtedly impact upon the accessibility of websites, both Lazar et al. (2004) and Farrelly (2011), in their respective models, consider web development tools, guidelines and resources to have an even greater influence. According to Lazar et al., "good, well-written guidelines and powerful software tools" are likely to improve levels of accessibility, whereas "poorly-written, confusing guidelines, and hard to use or unclear software tools" are likely to reduce them (p. 4). Numerous studies, both of web developers and of the tools and guidelines they use, support these assertions. These studies not only question the validity and reliability of tools and guidelines but also demonstrate how web developers find them confusing and difficult to use.

Reflecting on an accessibility evaluation of 50 homepages in the mid-Atlantic United States, Lazar, Beere, Greenidge and Nagappa (2003) describe the tools used in the evaluation as "flawed, inconsistent and requir[ing] large numbers of additional manual checks" (p. 1). However automated the tools may be, the authors observe, they still rely on the subjective (and potentially flawed) judgement of human expert evaluators (see Section 2.1.1). Lazar et al. criticise the tools' lack of transparency (with regards to how they evaluate webpages against each accessibility guideline), which leads to inconsistencies in their output. They also highlight the tools' slavish adherence to certain guidelines (such as the presence of alternative text on images, which many tools pass regardless of its accuracy), which prevents them from truly reflecting the user experience of people with disabilities.

Many of the surveys discussed in Section 2.4 examined respondents' attitudes towards accessibility tool and guidelines. The webmasters surveyed by Lazar et al. (2004) reportedly found existing accessibility tools to be flawed and argued that tool improvements would motivate them to create accessible websites. Web developers surveyed by Rosson et al. (2005) admitted omitting accessibility checks due to relatively tedious and time-consuming tool support. Existing accessibility evaluation tools, the authors claim, were too verbose, prone to reporting false positives, and lacked automation. Web developers surveyed as part of the BenToWeb project (Petrie, 2006) also cited the difficulties they encountered with tools and guidelines as a major barrier to web accessibility.

The difficult and time-consuming nature of accessibility evaluation tools was raised by the web developers surveyed by Trewin et al. (2010). They described existing tools as being unclear, cumbersome, and incomplete, with respect to the standards that must be met. Confidence in the tools' output was also a big issue for respondents, with many reporting to have struggled with tools that generate many false positives. Surprisingly few of the web practitioners interviewed by Farrelly (2011) claimed to have used accessibility evaluation tools but those that had raised similar concerns over false positives, false negatives and insufficient support to remediate problems.

The study of accessibility evaluation tools by Petrie et al. (2007) considered whether, irrespective of the tools' performance (about which the authors agree they have "serious questions" (p. 2)), the usability of the tools presented a further barrier to accessible web development. The authors point out that web developers should be able to concentrate on web developments tasks and not be distracted by tools that are difficult to use or interpret.

Petrie et al. conducted a group expert evaluation of five popular accessibility evaluation tools (Bobby, Cynthia Says, Deque Ramp, Site Valet, and WAVE), in which they collectively noted and rated usability problems. The evaluators discovered what they considered to be very obvious, serious and easily avoidable usability problems in each of the five tools. The majority of problems related to how the tools reported accessibility issues, which the authors note was "almost always unclear" (p. 9). The tools were also reportedly unclear about the scope of evaluation, both in terms of the number of webpages being evaluated, and the guidelines/level of conformance being adhered to.

According to Petrie et al., entry-level accessibility evaluation tools were not making the vital task of web accessibility evaluation easy for web developers. Furthermore, the tools were doing very little to enhance web developers' knowledge and understanding of accessibility issues. Trewin et al. (2010) also concluded that web developers need tools that enhance their understanding of accessibility through detailed explanations of problems.

A novel attempt to overcome problems with existing tools was devised by Bigham and Ladner (2007). They developed a scripting framework called *Accessmonkey* that allows web users, web researchers and web developers to collaboratively improve accessibility. It uses JavaScript scripts to dynamically reconfigure webpages to automatically address accessibility problems, for example, by generating text alternatives for images or making components keyboard accessible. Bigham and Ladner note that existing tools "rarely offer specific suggestions that developers can implement" (p. 1) and that web developers struggle to integrate them into their existing workflows. Accessmonkey can be integrated into existing tools and browsers and provides a means of gaining specific feedback and workable solutions to accessibility problems directly from web users. However, while Accessmonkey is a pragmatic and empowering solution for disabled web users, it cannot guarantee that web developers will adopt or understand the proposed website adaptations. Reducing web developers' responsibility to an approval and editing process may have efficiency benefits but at the cost of failing to advance their understanding of the problems that users encounter.

Another innovative approach by Bigham, Brudvik and Zang (2010) also shifts some of the responsibility for web accessibility onto web users. *Accessibility by Demonstration* (*ABD*) is an extension to the open source web-based screen reader WebAnywhere (Bigham, Prince & Ladner, 2008). *ABD* allows WebAnywhere users to retroactively record accessibility problems as soon as they occur as human-readable macros and to send the recordings to web developers. The recordings capture both the screen reader output and whatever keyboard commands the user has made. This is accompanied by visual highlighting to assist web developers in following along. *ABD* plays the user's recorded actions over the current version of the page, allowing web developers to replay the recording while making iterative improvements to the webpage. While this approach does not furnish web developers with specific solutions to web accessibility problems, it gives them the opportunity to experience the problem from the user's perspective: a vital step in developing their understanding of web accessibility and stimulating creativity in solving accessibility problems in future.

Criticism for failing to enhance web developers' understanding has been levelled not only at tools but also at web accessibility guidelines. Not long after the publication of WCAG 1.0, Sloan, Gregor, Rowan and Booth (2000) highlighted the considerable amount of understanding required of web developers to use the guidelines effectively.

An evaluation of WCAG 1.0 by Colwell and Petrie (2001) found that the structure and tone of the document made it difficult for web developers to navigate and interpret. Furthermore, they found that, in some cases, adhering to the guidelines resulted in webpages being more inaccessible. Respondents to Knight's (2003) survey of designers, information officers and accessibility advocates ranked the difficulty of navigating, interpreting and implementing WAI guidelines, as the third biggest barrier to web accessibility. One respondent described the guidelines as "opaque, very poorly organised, daunting, and in many cases unrealistic" (para. 15).

Web developers surveyed by the DRC (2004) reportedly found WCAG 1.0 useful but considered the lack of simple but authoritative guidance a barrier to web accessibility. Sloan, Kelly, Heath, Petrie, Hamilton, and Phipps (2006) collated the criticisms made of WCAG 1.0 by other authors, characterising the guidelines as subjective, ambiguous, unnecessarily complex, logically flawed, and difficult to understand. They also highlighted the distinct lack of empirical evidence of the majority of recommendations made in WCAG 1.0. Despite acknowledging the WAI's significant accomplishments in defining and promoting web accessibility, Farrelly (2011) considered the confusion and uncertainty caused by its guidelines to be a leading factor in halting the diffusion of web accessibility.

With the publication of WCAG 2.0 in 2008, the WAI sought to address many of the criticisms levelled at the previous version of the guidelines. In doing so, however, they appeared to create further difficulties for web developers. Tackling the technology-specific nature of WCAG 1.0, which had led to the guidelines becoming out-dated, WCAG 2.0 was designed to be technology-neutral by promoting general accessibility concepts over specific implementation details. Unfortunately, this resulted in the use of ambiguous notions such as "accessibility supported technologies", which Alonso, Fuertes, González and Martínez (2010a) claim only caused further confusion.

Kapsi, Vlachogiannis, Darzentas and Spyrou (2009) evaluated the usability of WCAG 1.0 and 2.0. They considered the effectiveness of both versions of the guidelines to be impeded by the use of complex language, complex document structures, and lengthy documentation. While they acknowledged that WCAG 2.0 has overcome the technology dependency of its predecessor, they claimed it was still "characterised by an exponential learning curve" (p. 4).

Although few of the web practitioners interviewed by Farrelly (2011) were sufficiently familiar with WCAG, of those that were, all of them found the guidelines difficult to use. They raised concerns over the length of WCAG 2.0, its lack of clarity, obtuse language, and convoluted organisation. Participants considered the guidelines difficult for even seasoned web developers to use effectively, let alone hobbyists or anyone new to the field.

A study by Alonso, Fuertes, González and Martínez (2010b) of 17 students taking part in a web accessibility course concluded that WCAG is "far from testable for beginners" (p. 8). The authors attributed this to: difficulties in comprehending the language used in the guidelines; a lack of knowledge that is required to correctly evaluate the guidelines; and a reluctance to spend a lot of effort evaluating the guidelines.

Criticisms over the subjective nature of WCAG 1.0 were addressed by attempting to make WCAG 2.0 more objectively testable. A study by Brajnik (2009) with 35 student web developers, however, found that, for many of the WCAG 2.0 guidelines, web developers were unable to come to an 80% level of agreement about whether a problem was present in a webpage. Agreement between evaluators of many WCAG 2.0 criteria was actually lower than the equivalent criteria in WCAG 1.0, suggesting the second version of the guidelines are more difficult to apply.

Like its predecessor, WCAG 2.0 also suffers from a lack of empirical evidence. Though Power, Petrie, Freire and Swallow (2011) propose an effective methodology for remotely evaluating WCAG 2.0 implementation techniques, few members of the web accessibility community have continued this empirical, evidence-based approach. The validity of WCAG 2.0 is also questioned in an empirical study by Power et al. (2012) of the problems encountered by blind users on the web. This demonstrated that only 50.4% of the problems were covered by WCAG 2.0 success criteria.

2.6. Conclusions

The literature review reported in this chapter has highlighted the puzzling disparity between the plentiful efforts to support the creation of accessible websites and the difficulty or reluctance of web developers and other stakeholders to implement the guidance.

Over the last twenty years there have been numerous initiatives to support, encourage and compel web developers and other stakeholders to create websites that are accessible and usable by the broadest range of users, including people with disabilities. These initiatives have provided web developers with access to a suite of technologies for creating sophisticated, accessible websites and supported them in doing so with a variety of tools and a well-established body of accessibility information. Yet, despite these well-intentioned initiatives and the wide availability of accessibility information, web developers still struggle to create accessible websites.

Web developers' difficulty or reluctance to create accessible websites does not appear to be born out of ignorance, nor by a lack of awareness or concern. Studies of web developers over the last twenty years indicate a growing awareness and social responsibility towards web accessibility. Studies instead tend to cite societal and organizational factors, such as a lack of education and training, difficulties in enforcing accessibility legislation, and client and organisational attitudes to web accessibility. The usability of accessibility tools and guidelines is also considered crucial to the development of accessible websites.

Such studies, however, have typically elicited data via large-scale online surveys. While this has resulted in greater numbers of respondents, it has often been at the expense of more detailed, qualitative information that would clarify the disparity between web accessibility awareness and implementation and mitigate the possibility of self-selection and social desirability bias. Few, if any, studies, examine web developers in-depth in the context of their working environments to actually determine why and how they struggle to create accessible websites. This highlights an apparent knowledge gap with regard to the role of web developers and their working practices. It also highlights an opportunity to gain a greater understanding of where accessibility support is currently lacking and how it can be improved.

Chapter 3. A systematic review of web accessibility evaluation literature

3.1. Introduction

A considerable number of studies over the past fifteen years have evaluated the accessibility of websites for people with disabilities. Such studies tend to compare the accessibility of a group of websites either at particular points in time (e.g. Sullivan & Matson, 2000; Lazar et al., 2003; Olalere & Lazar, 2011), or over specified periods of time (e.g. Hackett, Parmanto & Zeng, 2005; Lazar and Greenidge, 2006; Loiacono, Romano & McCoy, 2009). Researchers typically choose a selection of websites to evaluate. This is often based on the most popular websites at the time (according to rankings produced by companies such as Alexa or Google) but may be limited to a particular country or sector, such as government, banking, or education. Then, using one or more methods (including automatic testing, manual inspection and, sometimes, user testing with disabled people: see Section 2.1), researchers evaluate the websites against various accessibility guidelines, standards and heuristics to identify accessibility problems and determine a measure or benchmark of web accessibility.

Consistent application of evaluation methods and comprehensive reporting of outcomes should allow researchers to build up a picture of the progress and current state of web accessibility. This would provide an indication, for example, of whether governmental policy objectives on web accessibility are being satisfied in a particular country. It could also identify specific websites, sectors, regions or countries in which web accessibility is lacking. A consistent, comprehensive, methodological approach would also provide a clearer indication of the longitudinal impact of guidelines, tools and other initiatives aimed at improving web accessibility. The latter was the motivation for the current study, which presents a systematic review of studies that have evaluated web accessibility published over the 15-year period between 1999 and 2014[16].

The considerable methodological variability that this review uncovers, however, with studies varying on just about every aspect of web accessibility evaluation, emphasises the difficulty in drawing robust conclusion from this body of evidence. It also highlights

[16] I collected the publications for this review together with a fellow PhD student at the University of York, Andreas Savva. In the end, Andreas did not use the material for his thesis and the analysis is entirely my own.

the need for greater consensus among researchers over how they conduct and report web accessibility evaluation studies. This would go some way to ensure that the outcomes of such studies are reliable, comparable, relevant and useful to researchers, practitioners, developers, and other interested stakeholders. In lieu of such consensus, however, this study extracts and synthesises the usable, comparable data from 397 web accessibility evaluation publications with the aim of reviewing how empirical research on web accessibility is conducted, understanding the changing state of the art in the field, and determining whether the accessibility of websites has improved over the last two decades.

3.2. Method

A systematic literature review offers a means of gathering, analysing and interpreting all available research that is relevant to particular area of research (Kitchenham & Charters, 2007). By using a trustworthy, rigorous and auditable methodology, systematic literature reviews allow for a fair evaluation of a particular research topic. The systematic literature review process comprises several stages: *planning the review*, *conducting the review*, and *reporting the review* (Kitchenham & Charters, 2007). *Planning the review* involves identifying the need for a review, specifying the research questions, and developing a review protocol. *Conducting the review* involves identifying and selecting the primary research studies, defining the quality assessment used to include studies, extracting and monitoring the data, and synthesising the obtained data. Finally, *reporting the review* involves specifying the dissemination mechanisms, and formatting and evaluating the report. The follow sections describe the planning and conducting of my systematic literature review. Subsequent sections report the outcome.

3.2.1. Research question

The aim of this study is to explore the web accessibility evaluation literature to address the following research question: *Has the accessibility of websites improved over the last two decades?* This analysis draws upon a comprehensive sample of published empirical works in this field, covering a range of disciplines and approaches. The criteria used to identify relevant studies are outlined below.

3.2.2. Identifying and selecting the primary research studies

An initial search in Google Scholar for the term "web accessibility" indicated that relevant studies were spread across numerous scientific journals and conferences, representing a range of disciplines, including HCI, library science, leisure and tourism, and even accountancy. To achieve as broad coverage of the literature as possible, therefore, nine widely used publisher's portals and databases of predominantly peer-reviewed publications were identified.

- ACM Digital Library (ACM) (http://dl.acm.org)
- IEEE Xplore Digital Library (IEEE) (http://ieeexplore.ieee.org)
- Web of Science (WoS) (http://webofscience.com)
- The Collection of Computer Science Bibliographies (CCSB) (http://liinwww.ira.uka.de/bibliography)
- Scopus (http://www.scopus.com)
- Science Direct (SD) (http://www.sciencedirect.com)
- Emerald Insight (EI) (http://www.emeraldinsight.com)
- Wiley Online Library (Wiley) (http://onlinelibrary.wiley.com)
- Google Scholar (GS) (http://scholar.google.co.uk)

The following search terms (and word variants) for retrieving the studies were defined:

- "website accessibility"
- "web site accessibility"/"web-site accessibility"[17]
- "webpage accessibility"
- "web page accessibility"/"web-page accessibility"[17]
- "accessibility of websites"
- "accessibility of web sites"/"accessibility of web-sites"[17]
- "accessibility of webpages"
- "accessibility of web pages"/"accessibility of web-pages"[17]

Searches for the exact phrases above were conducted across the full document text and metadata (including title, keywords, abstract etc.). Wherever publication databases gave the opportunity, the searches were restricted to English language publications, and to

[17] Each publication database or search engine treated a hyphen (-) the same as a space ( ) and returned the same results for these search term variations.

books, magazines, journal articles and conference proceedings only. To include the earliest and most recent studies, searches were not restricted by year of publication.

3.2.2.1. Inclusion and exclusion criteria

Publications that met the following criteria were *included*.

- Publications that evaluate the accessibility of a website or set of websites (see definition of website below). The evaluation may be based on a) the *entire* website, including all available webpages, b) a representative sample of webpages from each website (see definition of webpage below), or c) the homepage from each website (see definition of homepage below).
- Publications that evaluate the accessibility of a website or set of websites using any evaluation method (this includes, but is not limited to, automatic evaluation, manual inspection, and user evaluation of web accessibility: see Section 2.1).
- Publications that evaluate the accessibility of website or set of websites using any automated or manual tools and utilities.
- Publications that evaluate the accessibility of a website or set of websites using any set of guidelines, standards or heuristics (this includes, but is not limited to, WCAG 1.0, WCAG 2.0, and Section 508: see Section 2.2).
- Publications published in the so-called "black literature", i.e. publications that have been peer-reviewed and formally published in books, journal articles or conference proceedings.

Publications that met the following criteria were *excluded*.

- Publications that evaluate the accessibility of anything smaller than a single webpage. This includes, but is not limited to, specific web components, such as links, images, forms etc.
- Publications that evaluate only a single aspect of web accessibility, for example, the readability of text content, the availability of alternative text on images, and the destination of links.
- Publications that evaluate the accessibility of a website or set of websites that have been purposely created to include known accessibility problems and are not publicly available.

- Publications published in the so-called "grey literature", i.e. publications that have not been peer-reviewed or formally published in books, journal articles or conference proceedings (e.g. reports, theses or dissertations, newsletters etc.).
- Publications not written in English.

Based on a review of various definitions and sources (e.g. Wikipedia; Dictionary.com; Mozilla Developer Network), for the purpose of this thesis, a website is defined as:

- A collection of related webpages regarded as a single entity (e.g. www.bbc.co.uk or www.coca-cola.com) or a web application that fits on a single web page and has similar functionality to a desktop software application (e.g. www.gmail.com or www.twitter.com);
- Devoted to a single topic or several closely related topics from any sector;
- Typically identified with a common domain name (e.g. www.news.com);
- Published on a web server and publicly available via the web (This does not include websites that have been created solely for the purpose of evaluation); and
- Professionally maintained by a web developer (This does not include content that has been generated by users and cannot be expected to be accessible, such as an email in a webmail application, a document in an editing tool, a module in a web-based virtual learning environment, or a MOOC hosted by a repository).

A webpage is defined as:

- A web document composed of plain text and multimedia content (e.g. images, audio, video etc.);
- Formatted using Hypertext Markup Language;
- Accessed and transported via the Hypertext Transfer Protocol (HTTP); and
- Rendered via a web browser (e.g. Mozilla Firefox, Google Chrome etc.).

A homepage is defined as:

- The opening webpage of a website, which typically contains hyperlinks to other webpages on the same or other website(s).

Both Andreas Savva, who assisted me in collecting the publications for this review, and I separately evaluated each potentially relevant publication to determine whether or not it should be included. We then cross-checked each other's allocation of publications. In the event of any disagreement, we checked the publication again together. If there was still a disagreement, we would have asked a third researcher to arbitrate (however, this did not occur). We performed this procedure twice, at different levels of analysis: first by considering only a study's title and abstract, and second by considering the study's full text. This two-stage procedure vastly streamlined the study selection process and enabled us to process a considerable number of studies.

3.2.3. Data extraction strategy

Authors may report multiple website accessibility evaluations within a single publication. For example, they may have evaluated a website or websites using multiple evaluation methods (such as automatic evaluation and manual inspection: see Section 2.1). They may have also used multiple guidelines (such as WCAG 2.0 and Section 508), metrics (such WAB and WAQM) or tools (such as Bobby and Cynthia Says). Authors who are interested in longitudinal differences in web accessibility may have conducted accessibility evaluations over multiple years, or multiple evaluations with a single year.

To conflate the reporting of the multiple accessibility evaluations reported within such publications would lose much of the detail in the data. For instance, evaluations that are conducted using different automated evaluation tools tend to generate different results. Automated evaluation tools also tend to generate different results from human evaluators using manual inspection methods, which allow them to examine websites in more detail and eliminate any false positives. Evaluations conducted over even a short space of time may also generate varying results due to changes and updates to the website content. Attempting to derive a single metric of web accessibility from a publication that reports multiple accessibility evaluations would be misleading. To therefore provide a more accurate reflection of the accessibility of websites over the 15-year period, data about each individual website accessibility evaluation was extracted from each study.

In accordance with the research question, the following data (where available) was extracted about each individual website accessibility evaluation from each study.

- Publication details (e.g. publication year; author(s); title; name of publication; DOI/ISBN/URL).
- Year in which the evaluation was conducted.
- Website sampling strategy (e.g. homepage only; homepage plus the first level of webpages; other sampling strategy).
- Number of websites and/or webpages evaluated. Some publications did not explicitly report the number of webpages evaluated but it is possible to infer this information from the reported sampling strategy. For example, if the authors evaluated two webpages from ten websites, the number of webpages evaluated would be 20.
- Country (or countries) in which the website(s) under evaluation are located (e.g. a specific country, such as the USA or the UK; international websites).
- Type(s) of website under evaluation (e.g. a specific type, such as government or education; general/unspecified type of website).
- Evaluation method used (e.g. automated evaluation; manual inspection; user evaluation; other method): see Section 2.1 for an explanation of relevant evaluation methods).
    - If an expert evaluation is involved, number of expert evaluators.
    - If a user evaluation is involved, number of disabled/non-disabled users.
- Guidelines used in the evaluation (e.g. WCAG 1.0; WCAG 2.0; Section 508) and, where appropriate, the priority level used (A; AA; AAA): see Section 2.2 for an explanation of relevant guidelines and priority levels.
- Metrics used in the evaluation (e.g. WAB; WAQM; UWEM) and, where appropriate, the priority level used (A; AA; AAA): see Section 2.1.1 for an explanation of relevant metrics.
- Tool(s) and utilities used in the evaluation (e.g. automated evaluation tools; manual inspection utilities; web browsers; devices; ATs): see Section 2.1.1 and 2.1.2 for an explanation of relevant tools and utilities.
- Evaluation results (e.g. number and percentage of websites passing a particular level of conformance; metric score; task completion rate; number of problems identified by users; other type of result).

A number of steps were taken specifically to avoid duplication in the reporting of the evaluation results.

- Where different publications reported the same evaluation (e.g. when data from an earlier study was included in a later longitudinal study), only the results from the earliest publication were included.
- Where a publication had conducted more than one evaluation on the same set of websites (e.g. using different automated evaluation tools) the mean result from the different evaluations was calculated and included.

3.2.4. Conducting the review

The search to identify primary research publications was conducted in September 2014. The initial searches returned 3,107 potentially relevant publications. Table 3.1 shows a breakdown of these results according to publication database and search term.

Table 3.1: Number of initial search results by publication database and search term

	Publication database									
Search term	***ACM***	***IEEE***	***WoS***	***CCSB***	***Scopus***	***SD***	***EI***	***Wiley***	***GS*** [18]	***Total***
website accessibility	182	7	40	36	420	100	12	38	100	***935***
web site accessibility/web-site accessibility	138	13	27	25	333	56	62	37	100	***791***
webpage accessibility	6	1	0	0	8	3	1	1	58	***78***
web page accessibility/web-page accessibility	79	0	6	6	136	27	13	9	100	***376***
accessibility of websites	52	2	15	11	35	20	5	19	100	***259***
accessibility of web sites/accessibility of web-sites	76	6	25	14	75	21	26	14	100	***357***
accessibility of webpages	3	0	1	0	0	0	0	0	18	***22***
accessibility of web pages/accessibility of web-pages	79	4	13	13	47	15	14	4	100	***289***
Total	***615***	***33***	***127***	***105***	***1054***	***242***	***133***	***122***	***676***	***3107***

[18] Google Scholar returned 94,000 results for the term “website accessibility” alone. Given the impracticality of processing so many results, only the first ten pages of results were checked (approximately 100 publications per search term).

The initial searches inevitably returned a large number of duplicate publications, both *within* and *between* the results from each publication database. This was due to the similarity of the search terms used in the initial searches, typographical differences in how the publications were presented (e.g. some publications were prefixed by chapter numbers in some databases but not others), as well as the overlapping contents of the publication databases.

To remove the duplicate publications *within* each publication database, their details were entered into a spreadsheet and sorted by title. Duplicate publications were manually checked in the corresponding publication database. If confirmed to be duplicate publications, they were removed from the search results. In total, 585 duplicate publications were removed across each of the publication databases, leaving a total of 2522 studies. Table 3.2 shows a breakdown of the number of publications removed and the number of publications remaining, according to each publication database.

Table 3.2: Number of duplicate publications removed and unique publications remaining by publication database

	Publication database									
	ACM	***IEEE***	***WoS***	***CCSB***	***Scopus***	***SD***	***EI***	***Wiley***	***GS***	***Total***
Number of initial search results	615	33	127	105	1054	242	133	122	676	***3107***
Number of duplicate publications removed	127	4	9	23	182	57	24	24	135	***585***
Number of unique publications remaining	488	29	118	82	872	185	109	98	541	***2522***

To remove the duplicate publications *between* each publication database, the search results from each publication database were merged and, again, sorted by title. Duplicate publications were manually checked in the corresponding publication database. If confirmed to be duplicate publications, they were removed from the search results. In total, 834 duplicate publications were removed across each of the publication databases, leaving a total of 1,688 unique and potentially relevant publications for more detailed evaluation.

After applying the inclusion/exclusion criteria documented above to each publication's title and abstract, a further 1,303 publications were removed, leaving a total of 385 unique and potentially relevant publications. After applying the inclusion/exclusion criteria to each remaining publication's full text, a further 122 publications were removed (4 of these publications were, despite best efforts, unavailable). This left a remainder of 263 unique and relevant publications.

3.2.5. Augmenting the review

To ensure complete coverage of all relevant publications, in January 2016 the publication record of each of the 415 authors of the 263 relevant publications identified in the previous stage was checked. This had the added benefit of ensuring that any publications from the last quarter of 2014 (after the initial searches in September 2014) would be included. Each of the 415 authors' publication record was checked in three locations.

- The author's personal website (where available).
- Google Scholar (GS) (http://scholar.google.co.uk).
- Computer Science Bibliography (DBLP) (http://dblp.uni-trier.de).

Once duplicate publications had been removed using the same process described in the previous stage, an additional 387 potentially relevant publications remained. After applying the same inclusion/exclusion criteria used in the previous stage, 256 publications were removed (14 of these publications were, despite best efforts, unavailable), leaving a remainder of 131 additional publications to add to the corpus.

A further 3 relevant publications were found serendipitously during the course of the systematic literature review. After applying the same inclusion/exclusion criteria used in the previous stage, these were added to the final corpus of publications.

Overall, of the 2,078 potentially relevant publications found via the initial searches of publication databases and Google Scholar (1,688), the follow-up searches of author publication records (387), and via serendipitous discovery (3), a total of 397 relevant research publications were identified for inclusion in this systematic review. References to these publications are found in Appendix G.

3.3. Results

3.3.1. Publication frequency

Figure 3.1 presents the publication frequency of the 397 publications selected for inclusion in this systematic review. It indicates a growth of interest in the topic between 1999 and 2014, which peaks around 2011, and begins to decline thereafter.

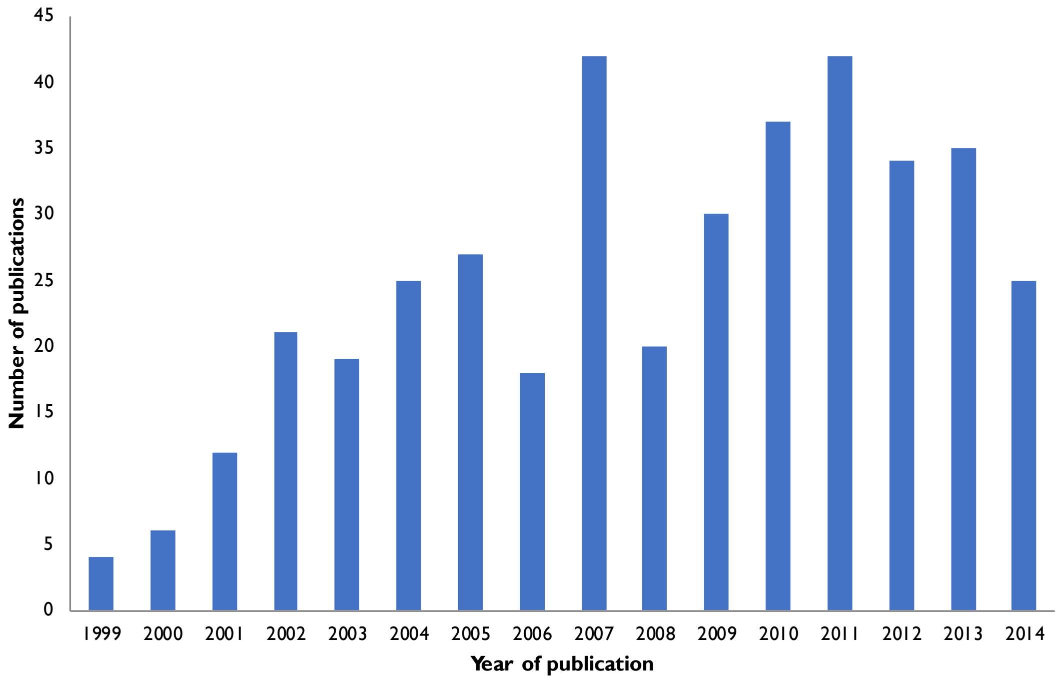


Figure 3.1: Publication frequency by year

(N = 397)

3.3.2. Evaluation frequency

As outlined in Section 3.2.3, authors typically report multiple web accessibility evaluations within a single publication. To avoid conflation of the results, data about each individual evaluation was extracted from each publication. Within the 397 publications from between 1999 and 2014, there were 755 individual evaluations, a mean of 1.9 evaluations per publication. The scope of each publication varies in terms of the number of years it covers: some publications include evaluations conducted in the same year as publication; many publications include evaluations conducted one, two or (in the case of longitudinal studies) over many years prior. Consequently, the period of time over which the evaluations were conducted (1997-2014) is slightly longer than the period of time over which they were published. Figure 3.2 presents the evaluation frequency of the 755 evaluations extracted for this systematic review.

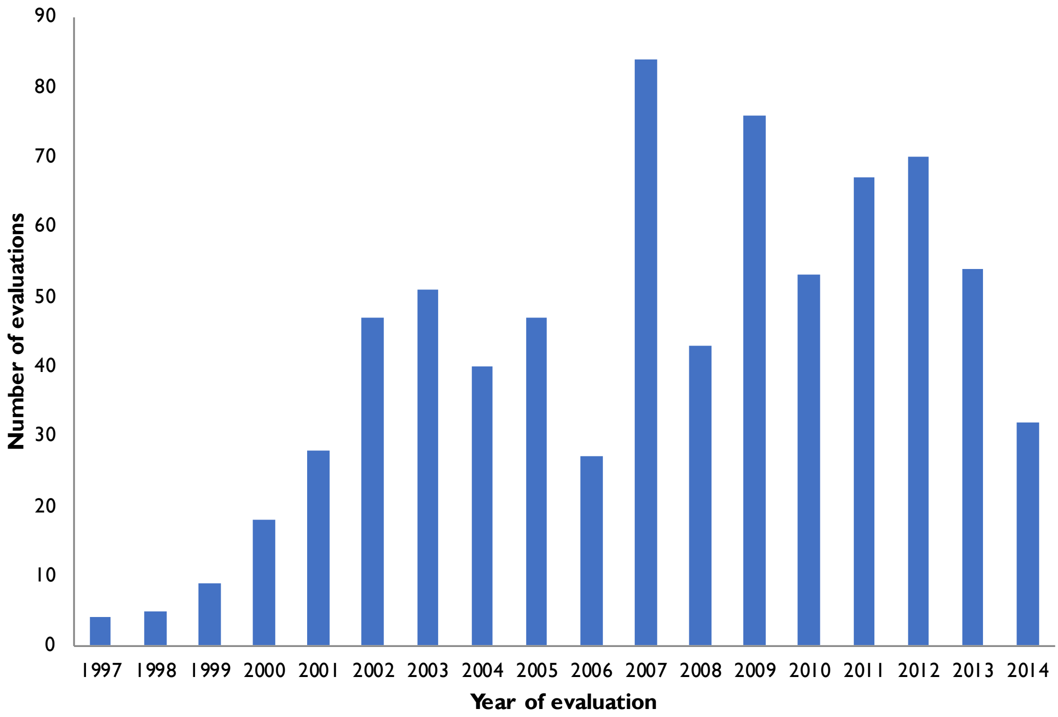


Figure 3.2: Evaluation frequency by year
(N = 755)

As with the publication frequency, this indicates a growth of interest in the topic, which peaks around 2007, and begins to decline thereafter. One explanation for this frequency pattern may be the publication of WCAG 2.0 in December 2008 (see Section 2.2). Given the substantial revision undertaken on version two of the guidelines, both in terms of their content and presentation, researchers and practitioners inevitably required some time to understand and interpret them correctly. Similarly, accessibility evaluation tool developers took several years to accommodate the new guidelines into their products. Another explanation for the recent decline in the number of evaluations may have been the rebranding and eventual discontinuation of the highly-popular automatic accessibility evaluation tool, Bobby. As described in Section 2.1.1, Bobby (or WebXACT, as it had been rebranded) ceased to be available as a free service and was taken offline in February 2008. Ultimately, the publication of a new and complex set of accessibility guidelines, coupled with a lack of tools and resources to support its application, may have contributed to the decline in the frequency of web accessibility evaluations and publications.

3.3.3. Website and webpage sampling

One of the many aspects by which web accessibility evaluations vary is the number and sampling of websites and the webpages they contain.

In some cases, particularly for smaller websites, it may be feasible to evaluate all webpages within a particular website. In many cases, however, where websites comprise hundreds, thousands, or even millions of webpages, evaluation of every webpage may not be possible or even desirable. Instead, evaluators typically select a representative sample of webpages from each website.

Web accessibility evaluation studies employ a number of strategies to identify the sample of webpages to evaluate. The most common strategy, used in 315 (42%) of the 755 evaluations, is to sample only the homepage of each website. In justifying this approach, studies frequently cite the early work of Nielsen (2000), in which he argues that the homepage represents a gateway to the rest of the website and sets the tone for the entire user experience. Yu (2002), Klein et al. (2003), and Loiacono and McCoy (2006) also consider the homepage to be representative of the rest of the website and the most likely access point for users. Hackett and Parmanto (2009), however, caution that the homepage may not be sufficient to determine the accessibility of an entire website. They argue that the accessibility of the homepage and all of the first-level pages linked from the homepage is more representative of the rest of the website.

Although extending the sample beyond the homepage is much less common, 53 (7%) evaluations include all first-level webpages linked from the homepage, 17 (2%) evaluations include all second-level webpages, and, 8 (1%) include all third-level webpages. In 69 (9%) evaluations, the webpage sample is based on a typical “user journey”, such as using an ecommerce checkout process, or navigating to a particular item of content. Another strategy, used in 93 (12%) evaluations, is to choose a random sample of webpages from each website. For 200 (27%) of the 755 evaluations, the authors do not report the webpage sampling strategy used. Table 3.3 presents the frequency and percentage distribution of evaluations by sampling strategy.

Table 3.3: Frequency/percentage distribution of evaluations by sampling strategy

Sampling Strategy	***Frequency***	***Percent***[19]
Homepage only	315	42
Homepage + 1st level webpages	53	7
Homepage + 2nd level webpages	17	2
Homepage + 3rd level webpages	8	1
User journey	69	9
Other sampling strategy	93	12
Not reported	200	27
Total	***755***	***100***

Some variation in the size of the sample of websites and webpages is inevitable. The authors of many of the studies included in this systematic review claim to have followed the recommendations provided in UWEM (Velleman et al., 2006) (see Section 2.1.2). This describes how the size of the sample of webpages necessary to evaluate a particular website depends on many factors, including: the size; age; complexity; and consistency of the website; the web developers' adherence to formal development processes; the required level of confidence in the results; and the availability of prior evaluation findings. However, even taking such factors into account, the number of websites and webpages evaluated varies considerably from one evaluation to another. Some publications include evaluations of only a single webpage of a single website; other publications report the use of web crawling and spidering software to sample as many webpages from as many websites as possible within a particular domain or country.

Within the 755 evaluations, the number of evaluated websites ranges from 1 to 18,096 (mean: 191; SD: 1103; mode: 1; median: 32). For 28 (4%) of the 755 evaluations, the authors do not report the precise number of evaluated websites. In such cases, they tend to report the number of webpages evaluated across all websites or refer only in vague terms to the website sample size. Table 3.4 presents the frequency and percentage distribution of evaluations by number of evaluated websites.

[19] Percentages throughout this chapter are rounded up to the nearest whole per cent.

Table 3.4: Frequency/percentage distribution of evaluations by number of websites evaluated

Number of websites evaluated	*Frequency*	*Percent*
1 website	87	12
2 to 5 websites	94	12
6 to 10 websites	70	9
11 to 20 websites	67	9
21 to 50 websites	150	20
51 to 100 websites	100	13
101 to 200 websites	75	10
201 to 500 websites	43	6
501 to 1000 websites	18	2
1001+ websites	23	3
Not reported	28	4
Total	***755***	***100***

The number of evaluated webpages also varies considerably from one evaluation to another. Within the 755 evaluations, it ranges from 1 to 28,135,102 (mean: 129,537; SD: 1,884,240; mode: 10; median: 75). For 310 (41%) of the 755 evaluations, the authors do not report the number of evaluated webpages. In such cases, they tend to report only the number of websites evaluated. Table 3.5 presents the frequency and percentage distribution of evaluations by number of evaluated webpages.

Table 3.5: Frequency/percentage distribution of evaluations by number of webpages evaluated

Number of webpages evaluated	*Frequency*	*Percent*
1 webpage	2	<1
2 to 5 webpages	20	3
6 to 10 webpages	42	6
11 to 20 webpages	30	4
21 to 50 webpages	93	12
51 to 100 webpages	94	12
101 to 200 webpages	60	8
201 to 500 webpages	31	4
501 to 1000 webpages	17	2
1001+ webpages	56	7
Not reported	310	41
Total	***755***	***100***

The number and sampling of websites and webpages within them has implications both for the type of evaluation method used and the validity and reliability of the results. Whereas authors are able to manually evaluate small samples of webpages, they must rely on automatic tools to both gather and evaluate larger samples. Not all aspects of web accessibility, however, can be determined automatically, and, as described in Section 2.1, many (such as whether a text alternative accurately describes an image) rely on human judgement. Evaluations that draw upon a large sample of websites and/or webpages must inevitably rely upon automated methods and can, therefore, only provide a limited assessment of web accessibility.

3.3.4. Website type and location

Another aspect by which web accessibility evaluation studies vary is the type of websites that authors choose to evaluate. Some studies evaluate only a specific type of website. This is particularly common in government and education sectors, where online access to information and services is increasingly vital. Some studies evaluate a range of website types, often to indicate the progress and current state of web accessibility in general or to identify specific areas in which web accessibility is lacking.

Table 3.6 presents the frequency and percentage distribution of evaluations by website type. The total is greater than the overall number of evaluations due to some studies including multiple website types in a single evaluation. It can be seen from the table that accessibility research is predominated by evaluations of governmental/non-profit organisations (NPOs) and educational/academic websites, each of which account for 22% of the 755 evaluations.

Web accessibility evaluation studies also vary according to the location of evaluated websites. Some studies evaluate websites located in a specific country. The purpose of this might be, for example, to explore that country's commitment to governmental policy objectives on web accessibility, or to provide a web accessibility benchmark that can be compared to that of other countries. Some studies evaluate websites across multiple countries. The motivation for this is often to provide an international indicator of the state of web accessibility or simply because the World Wide Web is inherently international. Some studies combine the two approaches by comparing the accessibility of websites in a specific country (often one that is developing) to international web accessibility levels.

Table 3.6: Frequency/percentage distribution of evaluations by website type

Website type	***Frequency***	***Percent***
Governmental/non-profit organisations (e.g. websites created by the local or national government of a country)	230	22
Educational/academic (e.g. websites that provide information about schools or universities)	229	22
General/unspecified	105	10
Business/commercial (e.g. websites that provide information about businesses, organisations or services)	94	9
Libraries/museums/archives (e.g. websites that provide information about academic or public libraries, museums, and archives)	78	7
Leisure/entertainment (e.g. websites that provide services such as streaming media)	71	7
Content generation (e.g. wikis, blog, content management systems, and web-based applications)	56	5
Ecommerce (e.g. websites offering goods and services for sale and enabling online transactions)	53	5
Health/medical (e.g. websites that provide health information)	53	5
News (e.g. websites that provide news, politics and commentary)	25	2
Search engines/directories (e.g. websites that index and curate material on the Internet)	22	2
Banking/finance (e.g. websites that provide financial information)	18	2
Social (e.g. websites that allow users to share and comment on stories, images, videos and other information)	14	1
***Total*[20]**	***1048***	***100***

The 755 evaluations evaluate websites located in 72 countries across 6 continents. For 215 (28%) of the 755 evaluations, the authors do not report a specific country or continent. Table 3.7 presents the frequency and percentage distribution of evaluations by continent. The total is greater than the overall number of evaluations due to some studies including websites located in multiple countries in a single evaluation. It can be seen from the table that accessibility research is predominated by evaluations of websites located in Europe and North America, each of which account for 24% of the 755 evaluations.

[20] The total is greater than the overall number of evaluations due to some studies including multiple website types in a single evaluation.

Table 3.7: Frequency/percentage distribution of evaluations by continent

Continent	*Frequency*	*Percent*
Europe (including Austria, Belgium, Croatia, Cyprus, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Romania, Serbia, Spain, Sweden, Switzerland, and the UK)	264	24
North America (including Canada and the USA)	261	24
Asia (including Bangladesh, Cambodia, China, Hong Kong, India, Indonesia, Iran, Israel, Japan, Jordan, Korea, Malaysia, Nepal, Oman, Pakistan, Palestine, Philippines, Saudi Arabia, Singapore, Sri Lanka, Taiwan, Thailand, Turkey, and the United Arab Emirates)	146	14
South America (including Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Guyana, Peru, Paraguay, Suriname, Uruguay, and Venezuela)	80	7
Africa (including Angola, Kenya, Liberia, Mozambique, Namibia, Nigeria, South Africa, Uganda, and Zimbabwe)	62	6
Australia/Oceania (including Australia and New Zealand)	50	5
International	215	20
***Total*[21]**	***1078***	***100***

3.3.5. Evaluation methods and tools

Web accessibility evaluation studies employ a number of methods to evaluate the accessibility of websites. As described in Section 2.1, these can be broadly categorised as automated evaluation, manual inspection, and user evaluation methods. Table 3.8 presents the frequency and percentage distribution of evaluations by method.

Table 3.8: Frequency/percentage distribution of evaluations by method

Method	*Frequency*	*Percent*
Automated evaluation	510	68
Manual inspection	189	25
User evaluation	56	7
Total	***755***	***100***

Over two thirds (510, 68%) of the 755 evaluations use automated evaluation methods to determine the accessibility of websites. Not all of these evaluations provide details of the specific automated tools used. Of the evaluations that do, however, Table 3.9 presents the six most frequently reported. These include commercial tools as well as tools that researchers and practitioners have developed themselves. Percentage distribution is given as a proportion of the overall number of evaluations (755). The

[21] The total is greater than the overall number of evaluations due to some studies including websites located in multiple countries in a single evaluation.

table clearly indicates the dominance of the automated evaluation tool, Bobby (later rebranded as WebXACT) by CAST/Watchfire, which is used in over a fifth (22%) of the evaluations.

Table 3.9: Most frequently used automated accessibility evaluation tools

Automated tools	***Frequency***	***Percent***
Bobby/WebXACT by CAST/Watchfire	169	22
TAW[22] by Fundación CTIC	39	5
WAVE[23] by WebAIM	27	4
Hera[24] by Fundación Sidar	23	3
AChecker[25] by the Inclusive Design Research Centre	21	3
Cynthia Says[26] by Cryptzone	21	3

A quarter (189, 25%) of the 755 evaluations use manual inspection methods to determine the accessibility of websites. Few of these evaluations provide details of the specific tools, utilities, web browsers and devices used. Of the evaluations that do, however, the most frequently reported tools and utilities include the Web Developer Toolbar[27] developed by Chris Pederick, (used in 7 (1%) evaluations), the Juicy Studio suite of tools[28] developed by Gez Lemon (used in 4 (1%) evaluations), and the Web Accessibility Toolbar (WAT)[29] developed by The Paciello Group (used in 3 (<1%) evaluations). The most frequently reported web browser is Microsoft Internet Explorer (used in 18 (2%) evaluations), followed by the Lynx text-only web browser (used in 13 (2%) evaluations) and Mozilla Firefox (used in 12 (2%) of evaluations).

Not all of the evaluations that indicate the use of ATs provide specific details about them. Of the evaluations that do, however, Table 3.10 presents the six most frequently reported ATs. Percentage distribution is given as a proportion of the overall number of evaluations (755). Though expert evaluators rely on a variety of ATs, the table indicates that the most common is the JAWS screen reader by Freedom Scientific, used in 5% of evaluations.

---

[22] https://www.tawdis.net/
[23] https://wave.webaim.org/
[24] http://www.sidar.org/hera/
[25] https://achecker.ca/
[26] http://www.cynthiasays.com/
[27] https://chrispederick.com/work/web-developer/
[28] http://juicystudio.com/services.php
[29] https://developer.paciellogroup.com/resources/wat/

Table 3.10: Most frequently used assistive technologies

Assistive technology	*Frequency*	*Percent*
JAWS[30] by Freedom Scientific (screen reader)	62	8
NVDA[31] by NV Access (screen reader)	15	2
WindowEyes[32] by GW Micro/Ai Squared (screen reader)	8	1
ZoomText[33] by GW Micro/Ai Squared (screen magnifier/reader)	5	1
VoiceOver[34] by Apple (screen reader)	4	1
Dragon NaturallySpeaking[35] by Nuance (speech recognition software)	4	1

Manual inspection methods sometimes rely on the judgement of more than one expert evaluators, who inspect the accessibility of websites, either individually or as part of a group. Not all of the studies that indicate the use of manual inspection methods provide details of the number of expert evaluators conducting each evaluation. Of the studies that do, however, the number of expert evaluators per evaluation ranges from 1 to 25 (mean: 4.8; SD: 5.60; median: 3; mode: 1).

Only 56 (7%) of the 755 evaluations use user evaluation methods to determine the accessibility of websites. Not all of these studies provide details of the number and makeup of participants. Of those that do however, the number of participants per evaluation ranges from 1 to 112 (mean: 16.4; SD: 18.52; median: 11; mode: 4).

In almost three quarters (41, 73%) of the 56 evaluations involving users, participants comprise exclusively disabled people. Of these, the majority (86%) involve blind people, 38% involve partially-sighted people, 20% involve people with cognitive disabilities, 16% involve people with physical disabilities, 10% involve people with hearing disabilities, and 6% involve people with multiple disabilities. In 9 (16%) user evaluations, participants comprise both disabled and non-disabled people. Surprisingly, considering the focus of web accessibility on supporting people with disabilities, participants in 6 (11%) user evaluations comprise exclusively non-disabled people.

[30] https://www.freedomscientific.com/Products/Blindness/JAWS
[31] https://www.nvaccess.org/
[32] http://www.gwmicro.com/
[33] https://www.zoomtext.com/products/zoomtext-magnifierreader/
[34] https://www.apple.com/uk/accessibility/mac/vision/
[35] https://www.nuance.com/en-gb/dragon.html

3.3.6. Accessibility guidelines, conformance levels, and metrics

Web accessibility evaluation studies, with the exception of those that involve user testing, tend to evaluate websites against one or more sets of accessibility guidelines.

The most commonly-used sets of guidelines are WCAG 1.0 and 2.0. Together, these are used in over two thirds (507, 67%) of the 755 evaluations. Section 508, despite legally applying only to US federal government websites, is the third most commonly-used set of guidelines, used in 66 (9%) evaluations. In 44 (6%) evaluations, authors use various other guidelines, standards and heuristics to determine the accessibility of websites. Some of these are developed in academia and are theory-based, whereas others are developed in industry and are derived from practical experience. For 6 (1%) of the 755 evaluations, the authors do not report the set of guidelines used and in 132 (17%) evaluations, guideline conformance does not apply. Table 3.11 presents the frequency and percentage distribution of evaluations by set of guidelines.

Table 3.11: Frequency/percentage distribution of evaluations by guidelines

Set of Guidelines	***Frequency***	***Percent***
WCAG 1.0	378	50
WCAG 2.0	129	17
Section 508	66	9
Other	44	6
Not reported	6	1
Not applicable	132	17
Total	***755***	***100***

The 'other' guidelines, used in 44 (6%) evaluations, vary according to their intended purpose. Some authors use guidelines that apply to specific website types. For example, Tatomir and Durrance (2010) published the *Tatomir Accessibility Checklist* to evaluate the accessibility of library websites. Some authors use guidelines that support specific user groups who may not be adequately supported by regular guidelines. For example, several authors (e.g. Becker, 2004; Hart & Chaparro, 2004; Hart, Chaparro & Holcomb, 2004) evaluate websites according to the *Making Your Web Site Senior Friendly* checklist, published by the US National Institute on Aging and National Library of Medicine (NIA & NLM, 2002).

Some authors derive their own guidelines from existing sets of guidelines. For example, Chen, Chen and Shao (2006) use their own set of guidelines based on WCAG 1.0. Greeff and Kotzé (2009) use the *Fujitsu Web Accessibility Guidelines*, which incorporate aspects of WCAG 1.0. Stewart, Narendra and Schmetzke (2005) use the *Ohio State University Web Accessibility Standards*, which are based on Section 508. Similarly, Lazar et al. (2013) use the *Maryland IT Non-Visual Access Guidelines*, which are also based on Section 508. DiLallo and Siegfried (2009) evaluate websites not against a set of guidelines but against a WAI document entitled *Quick Tips to Make Accessible Web Sites* (Henry & Popolizio, 2001).

Some authors use guidelines that are local to their own country or region. For example, several authors (e.g. Lee, Kim & Kim, 2007; Park, Lim & Lim, 2014), use KWCAG, which is a localised version of WCAG developed in Korea. Greeff and Kotzé (2009) use the German BITV directive (Barrierefreie Informationstechnik Verordnung (Barrier free Information Technology Ordinance)), which corresponds closely to WCAG 1.0. Gambino, Pirrone and Di Giorgio (2014) evaluate websites according to the Italian Stanca Act, which is closely aligned with WCAG 2.0.

Some authors combine other sets of guidelines. For example, Thompson, Burgstahler and Moore (2007, 2010) use their own set of guidelines that draw upon both WCAG 1.0 and Section 508. Several authors (e.g. Calvo, Iglesias & Moreno, 2012; Moreno, Calvo & Iglesias, 2011; Moreno, Iglesias, Calvo, Delgado & Zaragoza, 2012) use a combination of WCAG 2.0 and ATAG 2.0 to evaluate the accessibility of rich Internet applications.

There is some overlap in usage between the two versions of WCAG, with many studies using still using WCAG 1.0 long after publication of WCAG 2.0 in 2008. Reasons for this vary. Some authors struggled to get to grips with WCAG 2.0 and found WCAG 1.0 easier to apply. Some authors continued to use WCAG 1.0 to aid comparison, either between different web accessibility evaluation studies (e.g. Comeaux & Schmetzke, 2010), or between the two versions of WCAG (e.g. Kamoun, Al Mourad & Bataineh, 2013). Some authors (e.g. Gonçalves, Martins, Pereira, Oliveira & Ferreira, 2013) used both sets of guidelines (and also Section 508) for the sake of obtaining the most comprehensive results. Some authors (e.g. Nizar, Obedidat & Abu-Addose, 2013) wanted to rely on a specific evaluation tool (such as the automated tool, Bobby, which could only evaluate websites to WCAG 1.0 or Section 508); other authors (e.g. Yu &

Parmanto, 2011) wanted to rely on a specific metric (such as WAB, which is closely associated with WCAG 1.0). The slow incorporation of WCAG 2.0 into legislation prompted some authors (e.g. Correia, Cruz, Nunes, Martins, Gonçalves, Paredes & Martins, 2013) to continue to use WCAG 1.0. Some authors (e.g. Melo, Soares, Campos & Correia, 2013) do not even acknowledge the second version of the guidelines.

Even among studies that evaluate websites against a specific set of guidelines, the selection of individual guidelines within each set varies. As described in Section 2.1.1, WCAG 1.0 and 2.0 both categorise individual guidelines according to three discrete conformance levels: Level A (lowest), AA, and AAA (highest). These reflect the impact of different guidelines upon users and provide a measure or benchmark of a website's accessibility. Table 3.12 presents the frequency and percentage distribution of evaluations by WCAG priority level.

Table 3.12: Frequency/percentage distribution of evaluations by WCAG priority level

Priority Level	***Frequency***	***Percent***
Level A	106	14
Level AA	52	7
Level AAA	302	40
Not reported	47	6
Not applicable	248	33
Total	***755***	***100***

In 106 (14%) of the 755 evaluations, authors evaluate websites only against the lowest level of conformance, Level A. In only 52 evaluations (7%), authors evaluate websites only against Level AA. In 302 evaluations (40%), authors claim to evaluate websites against the highest level of conformance, Level AAA. For 47 of the 755 evaluations (6%), the authors do not report a specific priority level and in 248 (33%) evaluations, priority levels do not apply. In some cases, the intended priority level was inferred from the evaluation results but in many cases, it is impossible to determine.

It became clear that many authors misunderstood the cumulative nature of the WCAG conformance levels, whereby websites cannot conform to Level AA without satisfying all Level A and AA requirements, and, similarly, cannot conform to Level AAA without satisfying all Level A, AA, and AAA requirements. Authors commonly report websites as being Level AAA conformant despite also reporting numerous Level A or AA violations. Where the number of violations was available, the results were corrected during the data extraction process.

Some authors report the accessibility of websites according to a particular metric (see Section 2.1.1). Although metric-based evaluations make up only a small proportion (117, 15%) of the 755 evaluations, the set of metrics that authors draw upon is diverse. The most commonly used metric is the WAB (Web Accessibility Barriers) score (Parmanto and Zeng, 2005), used in 50 (7%) evaluations. The two second most commonly used metrics are the UWEM score, which is part of the Unified Web Evaluation Methodology (Velleman et al., 2006), and the WAQM (Web Accessibility Quantitative Metric) (Vigo et al., 2007), both used in 12 (2%) evaluations. The third most commonly used metric is the OAM (Overall Accessibility Metric) (Bailey & Burd, 2005), used in 5 (1%) evaluations. Notably, each of these metrics is totally automatic; that is, they are derived solely from the output of automated accessibility evaluation tools. Although accessibility metrics exist that are derived from both automated and manual evaluation data (such as SAMBA (Brajnik & Lomuscio, 2007)), they appear to be less widely used. In 38 (5%) evaluations, authors use various other metrics to report the accessibility of websites. Table 3.13 presents the frequency and percentage distribution of evaluations by metric. In addition to the diversity of metrics that authors use, and the varying attributes from which each metric is calculated, it was observed that authors often calculate metrics inconsistently or inaccurately, resulting in an even more confused picture of web accessibility.

Table 3.13: Frequency/percentage distribution of evaluations by metric

Metric	***Frequency***	***Percent***
Web Accessibility Barrier (WAB)	50	7
Unified Web Evaluation Methodology (UWEM)	12	2
Web Accessibility Quantitative Metric (WAQM)	12	2
Overall Accessibility Metric (OAM)	5	1
Other	38	5
Not applicable	638	85
Total	***755***	***100***

3.3.7. Presentation of results

As well as differing on many aspects of methodology, including how it is documented, web accessibility evaluation studies vary according to how they format and present the results.

The most common approach, used in 324 (43%) of the 755 evaluations is to report the number of websites *or webpages* that conform *or do not conform* to a specific set *or subset* of guidelines. Almost a fifth (117, 15%) of evaluations present the results according to one or more web accessibility metrics. In 105 (14%) evaluations, authors measure accessibility according to the number of guideline violations, either per webpage, per website or per group of websites, whereas in 84 (11%) evaluations, authors describe the violated guidelines. In 55 (7%) evaluations, the accessibility of evaluated websites is arbitrarily categorised as very/mildly/not at all accessible according to the presence or absence of certain accessibility features (such as descriptive headings and labels or alternative text on images).

Table 3.14: Frequency/percentage distribution of evaluations by type of results presented

Type of results presented	***Frequency***	***Percent***
Number of conformant/non-conformant websites /webpages	324	43
Metrics (e.g. the WAB (Web Accessibility Barriers) metric)	117	15
Number of guideline violations or barriers (per webpage, website, or group of websites)	105	14
Description of violated guidelines or barriers	84	11
Guideline/feature compliance rate	55	7
Commentary of encountered problems	32	4
Task completion rate/time	15	2
Subjective rating of accessibility	10	1
Number/type of problems encountered by users	7	1
Not reported	6	1
Total	***755***	***100***

Studies that involve expert evaluators using manual inspection methods, such as barrier walkthrough, also tend to report the type, frequency and severity of accessibility problems or barriers. In 32 (4%) evaluations, authors provide a written commentary of the accessibility problems that they or users encounter. 15 (2%) evaluations involving user testing provide task completion rates or times, whereas 7 (1%) evaluations report the number and type of accessibility problems that users encounter. 10 (1%) user testing studies report users' subjective assessments of accessibility, either in the form of a grade or rating or the outcome of surveys and questionnaires.

Table 3.14, above, presents the frequency and percentage distribution of evaluations by type of results presented. In 6 (1%) of the 755 evaluations, the authors do not provide any indication of how the website or websites perform, instead focusing on other aspects of the investigation, such as the performance of tools or metrics.

3.3.8. Evaluation results

The different methods that authors use to determine the accessibility of websites, as well as the varying ways in which authors format and present the results, provides limited scope for determining the changing state of the art in web accessibility.

For example, evaluations that report the number of guideline violations or barriers, or the number of accessibility features, tend to do so variably by webpage, by website or by group of websites, making them difficult to compare. Web accessibility metrics, such as WAB and WAQM, *should* provide a more reliable measurement of web accessibility. However, the diversity of metrics employed across the various evaluations, coupled with their often inconsistent and inaccurate application, reduces the longitudinal effectiveness of this approach. Descriptions of violated guidelines and barriers, or commentaries of encountered problems serve to highlight various accessibility issues but also provide little basis for comparison. Similarly, measures of task completion or time, subjective ratings of accessibility, and the number and type of problems that users encounter are difficult to compare across different websites or user groups. Consequently, the only relatively consistent measurement of web accessibility across the 755 evaluations is guideline conformance – that is, the proportion of websites that conform to a particular set of guidelines.

Of the 755 evaluations, 437 report the number of websites that conform to one of the three most commonly used sets of guidelines (WCAG 1.0, WCAG 2.0 and Section 508). This is either reported explicitly or inferred from other measures (e.g. evaluations with any number of guideline violations constitute a fail). Some of these evaluations test only a small number of websites, meaning the proportion of conformant websites is not very representative. To avoid this, evaluations of fewer than five websites were excluded from the analysis, leaving a remainder of 387 evaluations. Table 3.15 presents the frequency distribution of evaluations by year of evaluation and set of guidelines.

Table 3.15: Frequency distribution of evaluations by year of evaluation and set of guidelines

Year	*WCAG 1.0 (A)*	*WCAG 1.0 (AA)*	*WCAG 1.0 (AAA)*	*WCAG 2.0 (A)*	*WCAG 2.0 (AA)*	*WCAG 2.0 (AAA)*	*Section 508*	*Total*
1999	2	-	-	-	-	-	-	***2***
2000	9	2	2	-	-	-	-	***13***
2001	13	1	1	-	-	-	1	***16***
2002	20	9	8	-	-	-	3	***40***
2003	20	8	8	-	-	-	6	***42***
2004	14	8	6	-	-	-	2	***30***
2005	13	7	7	-	-	-	4	***31***
2006	9	3	3	-	-	-	1	***16***
2007	17	13	15	-	-	-	5	***50***
2008	14	8	7	2	1	1	-	***33***
2009	2	-	-	3	1	1	6	***13***
2010	7	4	4	4	5	4	3	***31***
2011	3	2	1	5	-	-	2	***13***
2012	5	4	3	5	3	1	6	***27***
2013	5	5	5	3	3	3	-	***24***
2014	2	-	-	2	1	1	-	***6***
Total	***155***	***74***	***70***	***24***	***14***	***11***	***39***	***387***

As can be seen from the table, the biggest (155, 40%) and most evenly distributed proportion of the 387 evaluations determine conformance to WCAG 1.0 Level A. The number and distribution of evaluations in the remaining proportions is not sufficient for further analysis.

Focusing specifically on the 155 WCAG 1.0 Level A evaluations, Table 3.16 presents the frequency and percentage distribution of evaluations by proportion of conformant websites in each evaluation. This shows that in over half (87, 56%) of the evaluations, the proportion of conformant websites is 20% or less, and in a third (51, 33%) of the evaluations, the proportion is 10% or less. In none of the evaluations does the proportion of conformant websites exceed 75%.

Table 3.16: Frequency/percentage distribution of evaluations by proportion of WCAG 1.0 Level A conformant websites in each evaluation

Proportion of WCAG 1.0 Level A conformant websites	***Frequency***	***Percent***
0 – 10%	51	33
11 – 20%	36	23
21 – 30%	30	19
31 – 40%	17	11
41 – 50%	7	5
51 – 60%	6	4
61 – 70%	5	3
71 – 80%	3	2
81 – 90%	0	0
91 – 100%	0	0
Total	***155***	***100***

The type and location of websites sampled in the 14 highest-scoring (51-80% conformance) evaluations is predominantly educational/academic (6 of the 14, 43%) and government/NPOs (4, 29%) from North America (8, 57%) and Europe (7, 50%). This is consistent with the general predominance of these websites types/locations in accessibility research (see Section 3.3.4). Notably, each of the 14 highest-scoring evaluations uses automated evaluation methods (14, 100%) and all but two of them (12, 86%) use the automated evaluation tool, Bobby, or its successor, WebXACT. While, again, this is consistent with the most frequently used evaluation method and tools in accessibility research (see Section 3.3.5), the relatively high proportion of conformant websites reported in these evaluations may also signify the risk of relying exclusively on automated evaluation tools (see Section 2.1.1).

The type and location of websites samples in the 26 lowest-scoring (0% conformance) evaluations is more varied. The most common type of website sampled is government/NPOs (8 of the 26, 31%). The most common website locations are Asia (9, 35%), International (7, 27%), and Europe (6, 23%). Automated evaluation methods still dominate the lowest-scoring evaluations (19, 73%), although the tool used is more varied, and includes WAVE, Cynthia Says, and AChecker. Manual inspection methods are used in 7 (27%) evaluations. The fact that none of the websites in these evaluations were conformant may be due to the greater accuracy of tools other than Bobby. Alternatively, it may reflect the greater discrimination offered by manual inspection methods (see Section 2.1.2).

Figure 3.3 presents the mean proportion of WCAG 1.0 Level A conformant websites by year. The numbers of WCAG 1.0 Level A evaluations conducted in 1999, 2009, 2011 and 2014 are very small, and would not be very representative of those years. Rather than exclude those evaluations from the analysis, they were merged into adjoining years.

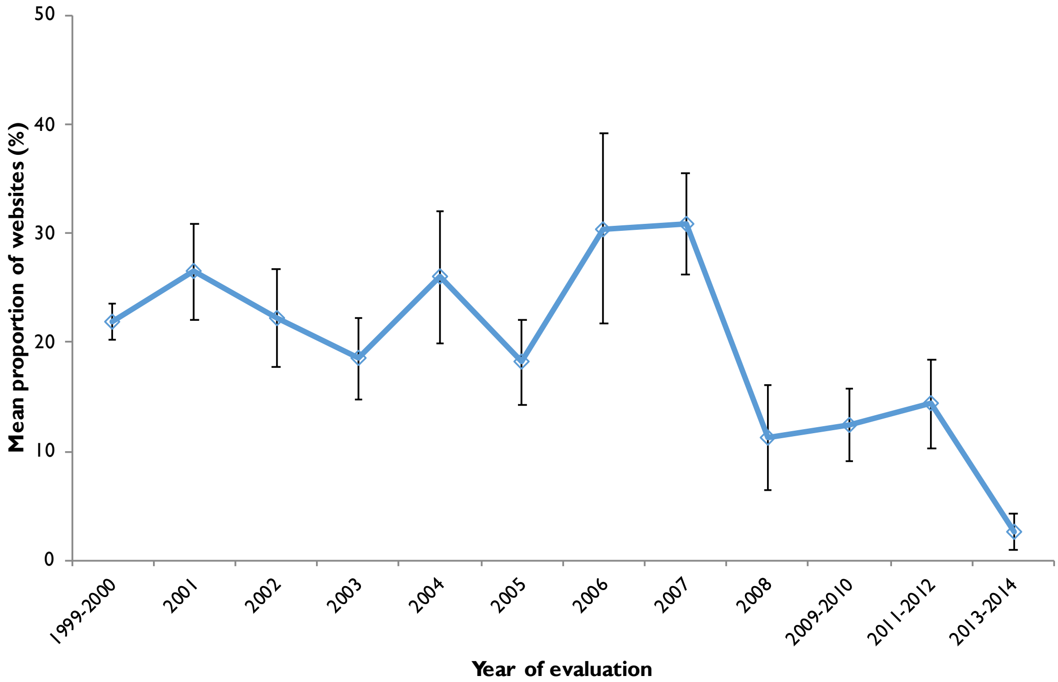


Figure 3.3: Mean proportion of WCAG 1.0 Level A conformant websites by year (N = 155; Error bars: standard error)

There are two observations to make from this figure. Firstly, the mean proportion of websites that conform to WCAG 1.0 Level A is extremely low over the 15-year period covered by this systematic review. It peaks barely above 30% at 30.84% in 2007 and plummets to 2.55% in 2013-2014. This indicates that between 1999 and 2014, on average, less than a third of evaluated websites conformed to the most basic level of web accessibility.

Secondly, levels of web accessibility conformance over the 15-year period covered by this systematic review show a persistent occurrence of accessibility problems that does not appear to be improving. To explore this further, Figure 3.4 presents a scatter plot and linear regression line of the proportion of conformant websites in each the 155 evaluations by year between 1999 and 2014. Again, evaluations conducted in 1999, 2009, 2011 and 2014 were merged into adjoining years but this is not reflected in the x-axis of the scatter plot, which includes each individual year.

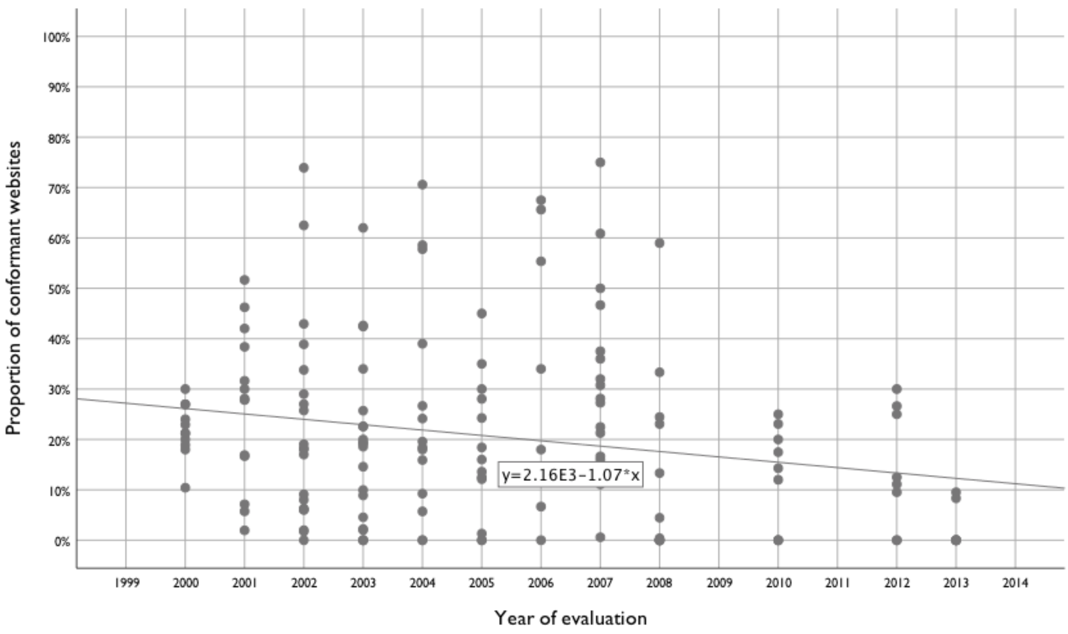


Figure 3.4: Scatter plot and linear regression line of the proportion of WCAG 1.0 Level A conformant websites in evaluations by year

(N = 155)

A simple linear regression was calculated to predict the *proportion of WCAG 1.0 Level A conformant websites* in each evaluation based on *year*. A significant regression equation was found, $F(1, 153) = 7.25$, $p = .008$, although this accounts for very little of the variance, $R^2 = .05$.

This suggests a modest but significant decline in the proportion of WCAG 1.0 Level A conformant websites over the 15-year period covered by this systematic review. However, given that the year of evaluation accounts for very little of the variance, there may be other factors that could not be factored out in the analysis. Many of the methodological inconsistencies observed in the previous sections that contribute to the overall variability and noise in the data, and which preclude a more precise analysis, could account for more of the variance observed in the evaluation results. These include the website and webpage sampling strategy (see 3.3.3), the type and location of the evaluated websites (see Section 3.3.4), and the choice of evaluation method and selection of tools (see Section 3.3.5). It is also important to acknowledge the proliferation and growing complexity of websites over the 15-year period, which was not assessed in this study, but which may have introduced further accessibility barriers and made conformance harder to achieve.

3.4. Discussion

The aim of this study was to review how empirical research on web accessibility is conducted, to understand the changing state of the art in web accessibility and to determine whether the accessibility of websites has improved over the last two decades. This involved a comprehensive and systematic review of website accessibility evaluation studies that have been published over a 15-year period between 1999 and 2014.

The reviewing process indicated an unanticipated degree of variability and noise in the data with studies varying on just about every aspect of web accessibility evaluation. Studies vary according to the number and sampling of websites and the webpages they contain; the type and location of the evaluated websites; the choice of evaluation methods; the selection of tools, browsers, devices and ATs; the selection of guidelines, standards and heuristics against which websites are evaluated; the number and sampling of user groups; the documentation of methodology; and the formatting and presentation of results. This considerable methodological variability made it extremely difficult to meta-analyse the data or draw reliable conclusions from it.

Nevertheless, despite this considerable noise, the usable, comparable data that could be extracted from 155 of the 397 publications selected for inclusion in this systematic review shows a persistent occurrence of accessibility problems that does not appear to be improving. Between 1999 and 2014, on average, less than a third of evaluated websites conformed to WCAG 1.0 Level A, the most basic level of web accessibility.

The type or location of the websites evaluated appears to have little bearing on accessibility levels. Different types of websites (e.g. governmental, educational, commercial etc.) across a range of countries are consistently found to have serious accessibility problems. Similarly, the choice of evaluation method and tools has had little impact on the results: evaluations using automated tools, evaluations relying on the judgement of one or more expert evaluators, and user evaluations with disabled people consistently report low levels of web accessibility.

These results are consistent with other, much smaller, longitudinal studies of web accessibility that cover much shorter and older time periods. For example, Hackett et al. (2005) evaluated the accessibility of a random selection of 40 websites over a five-year period from 1997 to 2002. Having access to the original source code allowed them to determine that the websites became progressively inaccessible while increasing in complexity. Loiacono et al. (2009) evaluated the accessibility of Fortune 100 websites between 2000 and 2005. They evaluated the homepage of each website against all three levels of WCAG 1.0 and established that, whilst the number of level A violations had decreased over time, the number of level AA and AAA violations had either remained the same or increased. Loiacono et al. surmise that an increased use of automated evaluation tools had allowed companies to correct the more obvious errors but at the expense of more thorough manual inspections. Hanson and Richards (2013) evaluated the accessibility of over 100 high-trafficked and government websites from the US and UK over a 14-year period from 1999 to 2012. They determined that despite some evidence of improvement (particularly for government websites), overall the websites exhibited generally low conformance with accessibility indicators. The authors speculate that any improvements in accessibility may be attributed to the development of good coding practices and SEO techniques rather than a growing awareness and adherence to WCAG guidelines.

To be able to properly compare web accessibility evaluation studies and to reliably benchmark the accessibility of websites, the accessibility community needs a consistent methodology with more complete methods of recording and reporting the data. In Europe, many member states established their own initiatives to analyse and monitor the accessibility of public sector websites, such as the Observatorio Español de Accesibilidad (Spanish Accessibility Observatory)[36] and the Polska Akademia Dostępności (Polish Academy of Accessibility)[37]. As part of a wider initiative to monitor the accessibility of Information and Communication Technology (ICT) products and services across Europe, the MeAC (Measuring Progress of eAccessibility in Europe) project evaluated websites from the 27 member states as well as 4 relevant third countries (Norway, Australia, Canada and the US). The most recent project report, MeAC III (Kubische, Cullen, Dolphin, Laurin, & Cederbom, 2013) indicated that there was still considerable room for improvement in web accessibility levels across Europe.

In response to the then proposed European Union Directive 2016/2102/EU on making websites and mobile apps of public sector bodies more accessible, a study led by the Swedish accessibility consultancy Funka proposed a unified European web accessibility monitoring methodology (Laurin et al., 2015). The methodology includes

[36] https://administracionelectronica.gob.es/pae_Home/pae_Estrategias/pae_Accesibilidad/pae_observatorio_accesibilidad_eng.html
[37] http://www.pad.widzialni.org

both automated and manual evaluation methods performed by experts, in conjunction with self-declaration by website owners and “end-user involvement” (essentially a feedback/complaint mechanism). It also recommends that the sample of pages be based on the recommendations made in WCAG-EM (Velleman et al., 2014). It is hoped that this will harmonise existing approaches to monitoring accessibility and make it much easier to measure and compare accessibility compliance across member states. While attempts to establish a unified accessibility monitoring methodology are laudable, they are currently limited to a specific geographic region (the EU) and to specific sectors within it (public sector websites). Extending the approach to incorporate a more global perspective and to monitor both public and private sector websites would better reflect the diversity of the web and provide a more accurate indication of the current state of web accessibility.

This systematic review is not without its limitations. According to Timmins and McCabe (2005), “the use of appropriate keywords is the cornerstone of an effective search” (p. 4). Given the focus of this systematic review, the initial search terms included variations on the phrase “web accessibility”. Although these search terms seemed appropriate, they returned not only a high number of publications to process but also a considerable number of irrelevant publications. Some publications were certainly relevant to web accessibility but did not feature a website evaluation. Many publications used the term “web accessibility” to refer to the performance or availability of websites rather than their accessibility to people with disabilities. Other publications had little to no relevance but had been returned due to the inclusion of a web accessibility statement on the website hosting the publication. It is unclear how this limitation could have been avoided: increasing the specificity of the search terms (to include, for example, “accessibility testing” or “accessibility guidelines”) may have increased the relevance of the results but at the cost of excluding publications that do not include such terms. An inspection of the keywords chosen by authors of web accessibility evaluation studies also confirms the appropriateness of the search terms: though some include additional keywords, such as “user testing” or “people with disabilities”, the majority of publications use one or more of the initial search terms.

Despite the use of appropriate search terms, additional checks of each author’s publication record revealed that the initial searches had missed a considerable number of publications (387 potentially relevant publications, later narrowed down to 131 relevant publications). The reasons for this disparity are unclear. One possibility is that the publications chosen by authors to publish their work were not indexed by any of the databases or publisher's portals examined. This seems unlikely, however, given the breadth of sources initially examined, which returned publications from across a range of disciplines. In addition to searching databases and publisher's portals, it may have been prudent to manually examine the proceedings of specific journals and conferences, such as Transactions on Accessible Computing (TACCESS) or the Web for All Conference (W4A). Whether this would have returned any additional publications, however, is questionable, as the initial search results included several publications from such sources. Another possibility is that, at the time the initial searches were conducted (September 2014), publications had not yet been indexed by any of the databases or publisher's portals. However, this seems only likely to apply to publications from the latter half of 2014, of which few emerged from the additional checks of each author's publication record. The additional publications were published as far back as 1999, which suggests another cause for the disparity. Of course, another possibility is inconsistencies in the selection and extraction of publications. However, attempts were made to cross-check the data during the study selection process.

Publication and selection biases pose a potential threat to the validity of systematic reviews. Publication bias is the tendency for only positive results to be published, whereas selection bias occurs when the sample of studies selected for analysis is not representative of the larger population of studies. Although no explicit attempt was made to control for publication bias in this study, no evidence of it was seen in the publications examined, which included both positive and negative results. Furthermore, there do not appear to be any financial, political or ideological interests that researchers or publishers of web accessibility studies may have for presenting a particular type of result. The inclusion of only peer-reviewed literature and exclusion of so-called "grey literature" and unpublished results may have resulted in selection bias and potentially threatened the generalisability of the work. For instance, a global report on web accessibility commissioned by the United Nations and conducted by the British accessibility company, Nomensa, was considered to be grey literature and therefore did not meet the inclusion criteria for this systematic review. The United Nations report was, however, further developed by Thompson, Burgstahler, Moore, Gunderson and Hoyt (2007), which *did* meet the inclusion criteria for this review. Nevertheless, this threat was alleviated by the selection of high quality research from a broad range of sources. Overall, the selection of publications in this systematic review is appropriate and representative of the web accessibility evaluation studies published, and these studies present a balanced overview of the state of web accessibility.

3.5. Conclusions

This chapter has presented a systematic review of studies that have evaluated web accessibility published over the 15-year period between 1999 and 2014. As established in Chapter 2, numerous initiatives over this period have sought to support, encourage and compel web developers to fulfil their responsibility to develop accessible websites. These include the formation of projects, working groups, and task forces; the definition of web accessibility standards, policies and legal imperatives; and the development of tools, guidelines and other resources. One would have expected such initiatives to herald an increase in levels of web accessibility, either through promoting greater knowledge and awareness of the topic or by creating a legal and moral imperative. Instead, the usable, comparable data that could be extracted from the 397 studies selected for inclusion in this systematic review shows a persistent occurrence of accessibility problems that does not appear to be improving. Between 1999 and 2014, on average, less than a third of evaluated websites conformed to WCAG 1.0 Level A, the most basic level of web accessibility.

The reasons for this surprising outcome are unclear. One explanation could be the sampling of websites used in the publications selected for review. Without a deeper analysis of the sampling strategies taken, it is impossible to determine the real-world applicability of the results. Even if this were possible, the proliferation and growing complexity of websites over the 15-year period may have simply outstripped the rate at which web developer knowledge, awareness and implementation of web accessibility has increased. Furthermore, it is impossible to determine whether developers of the evaluated websites were even exposed to WCAG and other web accessibility initiatives. More consistent methodology and complete methods of recording and reporting the data in web accessibility evaluation studies would have allowed a much deeper analysis, which may have resulted in a more reliable conclusion. Nevertheless, the outcome of this systematic review supports the notion that existing tools, guidelines and resources do not adequately support web developers and highlights an important research gap to be filled.

Chapter 4. A contextual inquiry of web developer working practices

4.1. Introduction

Despite finding considerable variability in how web accessibility evaluation studies are conducted and reported, the previous study showed a persistent occurrence of accessibility problems over the 15-year period between 1999 and 2014 that does not appear to be improving. Despite numerous initiatives and the wide availability of accessibility information, web developers are still failing to create accessible websites.

Existing attempts to explain the poor state of web accessibility rarely attribute it to a lack of awareness or concern on the part of web developers (e.g. Lazar, et al., 2004; Farrelly, 2011). Numerous surveys of web developers over the last two decades in fact indicate a growing awareness, both of web accessibility and of associated guidelines, tools and legislation. Studies instead tend to cite societal and organisational factors, such as a lack of training and education in web accessibility (e.g. Freire et al., 2008) or difficulties in enforcing accessibility legislation (Olalere & Lazar, 2011). Some studies (e.g. Putnam et al., 2012) call attention to negligent client and organisational attitudes to web accessibility, whereas others (e.g. Petrie et al., 2007; Brajnik, 2009) highlight difficulties in using accessibility tools, guidelines and resources. Few, if any, studies, however, examine web developers in-depth in the context of their working environments to actually determine why and how they struggle to create accessible websites. Furthermore, studies of web developers typically collect data via large-scale surveys that may encourage respondents to provide more socially acceptable responses and to portray themselves in a more positive light.

This chapter presents a contextual inquiry investigation into the working practices of 13 professional web developers. The objective of this study is to examine web developers in context to gain a greater understanding of their current working practices and to establish how web accessibility may be integrated into their existing workflows. It draws upon the ethnographically derived methodology of contextual inquiry (Beyer and Holtzblatt, 1997) to interview and observe web developers carrying out their own work in their own work environment.

4.2. Method

4.2.1. Design

This study uses a contextual design methodology to investigate the working practices of professional web developers. Contextual design is a user-centred design process established by Beyer and Holtzblatt (1997). It incorporates ethnographic methods, such as field observations and interviews, for collecting and analysing data about users and proposing design solutions to meet their needs. At the heart of contextual design is the contextual inquiry. This is a technique for observing and interviewing people as they carry out their work in the context of their own working environments. The data generated by this technique may then be formalised in a series of detailed models representing different aspects of the work under investigation. The ultimate goal of contextual design is to gain a greater understanding of how and why something is done (or why something is not done) and to identify how it may be improved.

The largely solitary nature of web development makes it a difficult type of work to investigate. It frequently involves internal mental processes and is largely conducted on a computer via a user interface. This can make it difficult for interviewers to observe the minutiae of the work being undertaken. Web development also requires a great deal of focus and concentration over long periods of time. Consequently, any observation of working practice—however discreet—may be disruptive to engaged web developers. Furthermore, whilst some aspects of web development (e.g. writing code) may be routine, other aspects (e.g. accessibility testing) may only occur intermittently, if at all. This intermittency could reduce the likelihood of contextual interviewers observing the task within the limited time available. Given these various challenges, I adopted a more interruptive approach to contextual inquiry in this study. This involved prompting web developers to carry out certain tasks (e.g. previewing a webpage) rather than waiting for those tasks to naturally occur. Though slightly more artificial, this more directed approach to contextual inquiry (which Beyer and Holtzblatt (1997) encourage in these situations, see pp. 73-76) ensured minimal disruption to the participant's working day and allowed me to focus on the aspects of their work that were most relevant to my investigation.

4.2.2. Participants

Thirteen professional web developers took part in the study. Unfortunately, in spite of efforts to recruit both male and female web developers, all the participants were male. Five web developers were aged 21-30 and eight were aged 31-40. Seven web developers were from the UK, three were from Ireland, and three were from Italy. Nine participants worked for i2Web consortium organisations. The others were recruited using opportunistic sampling from web developer mailing lists, online forums, social media, as well as personal contacts in the industry. The participants had between 1 and 15 years of experience of web development, with an average of 9 years. Six participants worked for large enterprises (250+ employees), six worked for SMEs (< 250 employees) and one participant was a self-employed freelancer. Participants were offered £30 worth of Amazon vouchers in compensation for their time and effort.

Though participants each identified as web developers, their specific roles and responsibilities varied. Five participants were exclusively involved in front-end web development. Their roles involved developing (and often designing) the user interface of websites, using client-side technologies such as HTML, CSS, and JavaScript. Four participants were exclusively involved in back-end web development. Their roles involve developing and maintaining the underlying architecture of websites, using server-side technologies such as PHP, Ruby, Python, Java, SQL and .Net. A further four participants were involved in both back-end and front-end web development (known as "full-stack development"). Their roles involve working cross-functionally on the full "stack" of web technologies, incorporating all aspects of web development. Four participants, in addition to developing websites themselves, were also responsible for managing development teams.

Table 4.1 summarises each participant's job title, organisation, sector and responsibilities.

Table 4.1: Job title, organisation, and responsibilities of participants

ID	*Job title*	*Organisation*	*Sector*	*Responsibilities*
WD01	Software Engineer	Large enterprise	Tele-communications	Designing and developing client-side and server-side software
WD02	Technical Manager	SME	Local government	Developing websites and administering servers
WD03	Web Developer	SME	Local government	Developing client-side and server-side software (specialising in security)
WD04	Development Manager	SME	Local government	Delivering software development projects and generating requirements
WD05	Web Developer/IT Support Specialist with Assistive Technologies	Large enterprise	Higher education	Developing websites/Providing IT support to staff and students, including updating and maintaining CMS-based websites
WD06	Junior Web Developer	Large enterprise	Pharmaceuticals	Preparing, updating and uploading web documents to the main corporate website
WD07	Web Designer	Self-employed	Various	Designing and developing websites; eliciting requirements; testing and evaluating websites; managing and coordinating projects; liaising with clients and stakeholders
WD08	Technical Project Manager/Software Architect/.Net Developer	SME	Information technology	Managing software development projects/Designing and developing software
WD09	Software Architect	SME	Information technology	Designing and developing software and hardware
WD10	Junior Web Developer	SME	Information technology	Developing client-side and server-side software
WD11	Solution Architect/Software Engineer	Large enterprise	Information technology	Working on EU-funded research projects to develop innovative software solutions.
WD12	Project Leader/Software Engineer	Large enterprise	Digital media	Managing a team of web developers/Developing websites
WD13	Software Engineer	Large enterprise	Digital media	Developing websites and apps for mobile and TV devices (specialising in multimedia software and video streaming)

4.2.3. Procedure

The contextual inquiries were conducted at each participant's place of work, sitting alongside them at their desk or workstation. The majority of participants worked in an open-plan office with their colleagues close by. One participant worked alone in his own office. The freelance web developer was interviewed in his regular work environment: a Starbucks café.

Each session was recorded using a wired USB webcam with microphone, which was positioned unobtrusively towards the edge of the participant's workspace, out of their direct sight. As well as capturing the participant's movements and speech, the webcam could be moved closer to anything of particular interest, such as something on the participant's screen or a physical artefact on the participant's desk. While this configuration was more challenging in the Starbuck's café, this was addressed by discreetly positioning the webcam much closer to the participant's laptop, out of direct sight of both the participant and other café patrons.

In accordance with the contextual design methodology proposed by Beyer and Holtzblatt (1997), each contextual inquiry had four parts:

Phase 1: The conventional interview. Following a brief introduction to the investigation and its aims, the participant was familiarised with the contextual inquiry methodology. They were assured that the information they would be providing would be confidential and anonymous and their consent was gained to take audio and video recordings for later transcription. Participants were then asked to read and sign an informed consent form (see Appendix A). This phase lasted approximately 15 minutes.

This phase of the contextual inquiry was also used to gain an overview of the participant's role, the organisation they work for, and the type of websites they typically develop. Although each participant was willing to take part in the contextual inquiry, some seemed a little uncomfortable with conducting this phase at their workstation, in close proximity to their colleagues. The participants' anxiety about being overheard by their colleagues raised concerns that they might provide socially acceptable answers to certain questions, particularly those related to web accessibility. For this reason, this phase of the contextual inquiry was conducted, where possible, in a private office, away from the participant's colleagues.

Phase 2: The transition. In this phase of the contextual inquiry, the focus shifted to the participant's actual desk or workstation, signalling the start of the contextual interview. The shift in approach was difficult to convey to some participants, who continued to treat it as a conventional interview. Nevertheless, following a brief re-statement of the contextual inquiry methodology, the participants soon adapted to the new approach. This phase lasted approximately 1-2 minutes.

Phase 3: The contextual interview. Participants started doing their work, whilst their activity was observed and interpreted. During this phase, participants were prompted to carry out certain tasks related to the development of websites. These included: previewing a webpage, validating a webpage, and testing the accessibility of a webpage. (The contextual interview schedule can be found in Appendix B). Participants were encouraged to carry out these tasks using the development tools they typically use and, where possible, working on a current or existing webpage they had developed themselves. During this time, the participants were discreetly observed, and asked questions based on what they were working on. Detailed notes were taken. Though the contextual design methodology encourages interviewers to follow participants around their workplace and monitor any interactions with other people, the participants in this study rarely moved from their desk or workstation and seldom communicated with others during the course of the interview (though they described when and with whom they might communicate during the course of their work in general). This phase lasted approximately 1.5 to 2 hours.

Phase 4: The wrap-up. Having brought the contextual interview phase to a close, a summary of the of the participant's work was given. This gave participants the chance to correct and elaborate upon the interpretation. Finally, participants were thanked for their time, offered a gift voucher, and asked whether they would be interested in taking part in further research. This phase lasted approximately 15 minutes.

The contextual inquiry methodology was piloted on a web developer who is a friend of the author, with the procedure refined in light of that experience.

4.2.4. Data preparation

In addition to taking detailed notes during the contextual inquiries, the audio and video recordings from each participant were transcribed and summarised. This generated a considerable amount of data. Sample transcripts from two participants can be found in Appendix G).

Having discussed the initial results with colleagues on the i2Web project, the key findings were incorporated into the i2Web project deliverable, *D3.2: Requirements for web developers and web commissioners in ubiquitous Web 2.0 design and development* (Petrie, Power & Swallow, 2011). The focus of this deliverable was the web development tools that are currently used by web developers and how web accessibility might be incorporated into those tools.

A more comprehensive analysis of the data was then conducted in accordance with the contextual design methodology proposed by Beyer and Holtzblatt (1997). This involved creating a series of work models from the perspective of each web developer, to represent what Beyer and Holtzblatt call the "five faces of work".

- **The flow model**, which represents the various people involved in the work, the interweaving of their roles and responsibilities, and the flow of information that passes between them.
- **The cultural model**, which represents the influences and pressures present in the work environment and the impact they have upon people's expectations, desires, values, and relationships.
- **The physical model**, which represents the physical environment in which work takes place, incorporating the desks and workstations, rooms, floors, and buildings in which people move about and work.
- **The sequence model**, which represents the sequence of activities that form work tasks, the intent behind each sequence, and the triggers that initiate them.
- **The artefact model**, which represents the (typically physical) objects that people create, use and modify in the course of doing work.

In each of the five work models, problems (known as "breakdowns") are represented by red lightning bolts.

Having created a set of work models for each web developer, these were merged to form a set of consolidated work models. The consolidated work models together represent an aggregated snapshot of the web developer sample in this study, revealing common patterns, structures and breakdowns in their work. The consolidated work models were then used to understand why web developers are struggling to make websites accessible and to identify how best to support them.

Aside from the cursory member checking that occurs during the "wrap-up" phase, the contextual design methodology proposed by Beyer and Holtzblatt (1997) does not suggest a means of validating and verifying the results. This is possibly because of its heavy emphasis on "building a shared understanding" through cooperative teamwork, which, to some extent, serves this purpose. Conscious of being a sole researcher, and the threat to validity and objectivity that this can pose, I sought to validate my analysis with an independent researcher. A common method of achieving this is through inter-coder reliability, which determines the extent to which independent judges make the same coding decisions and reach the same conclusions (Lombard, Snyder-Duch and Bracken, 2002). Due to the holistic nature of contextual inquiry, however, which requires researchers to form a deep and extensive understanding of participants' working practices, such an approach is neither practicable nor desirable.

Instead, I developed a series of ten questions relating to each of the different contextual design work models. For example, for the flow model, I asked: "Who does the web developer typically interact with?" For the physical model, I asked: "Where does the web developer's work typically take place?" (The full list of questions can be found in Appendix G). I then invited an independent researcher to answer the questions by reading the transcripts of two of the contextual inquiries (a subset of approximately 10% of the collected data). The researcher was familiar neither with web developers nor the contextual design methodology but had been trained in qualitative research methods. Although the transcripts would not give the researcher the same insights that being in the participants' presence had given me, they would still convey the information relevant to my investigation. A degree of consistency between the researcher's and my responses to these questions would then serve to corroborate the results.

With the exception of several minor details on which the independent researcher and I differed (and which we resolved through a negotiated agreement process), our responses were largely consistent. This process not only enhances the methodological rigour of my research, but also demonstrates a practicable and effective method for verifying and validating contextual inquiry data.

4.3. Results

The following sections describe in detail each of the five contextual design work models that were developed and include relevant quotes from the web developers.

4.3.1. Flow model

The flow model (see Figure 4.1) represents the various people involved in the work, the interweaving of their roles and responsibilities, and the flow of information that passes between them.

Despite the fairly solitary nature of web development, I discovered that few web developers work truly in isolation and most encounter a broad network of stakeholders. The web developers in this study typically interact with four distinct groups of individuals: colleagues and peers; immediate and more senior managers; clients; and in-house experts. None of the web developers in this study interact with website users.

4.3.1.1. Colleagues and peers

Except for the freelance web developer (who works alone), participants in this study frequently interact with their immediate colleagues (mentioned by 11 participants) or with web developer peers in other organisations (mentioned by 8 participants).

In most cases, the purpose of this interaction is to request or share technical help and advice. Interaction with immediate colleagues usually occurs face-to-face, unless someone is busy, when it might be conducted via email or instant message (IM) instead. Interaction with peers in other organisations is almost always conducted electronically: via email, instant message, IRC, mailing lists or fora, such as Stack Overflow[38] or The SitePoint Forums[39], which are virtual communities for programmers of any discipline to share knowledge and learn from others.

Several participants feel particularly grateful for the support and encouragement they receive from their colleagues and peers. One participant described his colleagues and peers as “very supportive, a proper community” and added “there is no spirit of ‘it’s mine, I’m having it’” $_{[WD5]}$. Because of this apparent collegiality, few breakdowns occur between web developers and their colleagues and peers.

[38] https://stackoverflow.com
[39] https://www.sitepoint.com/community

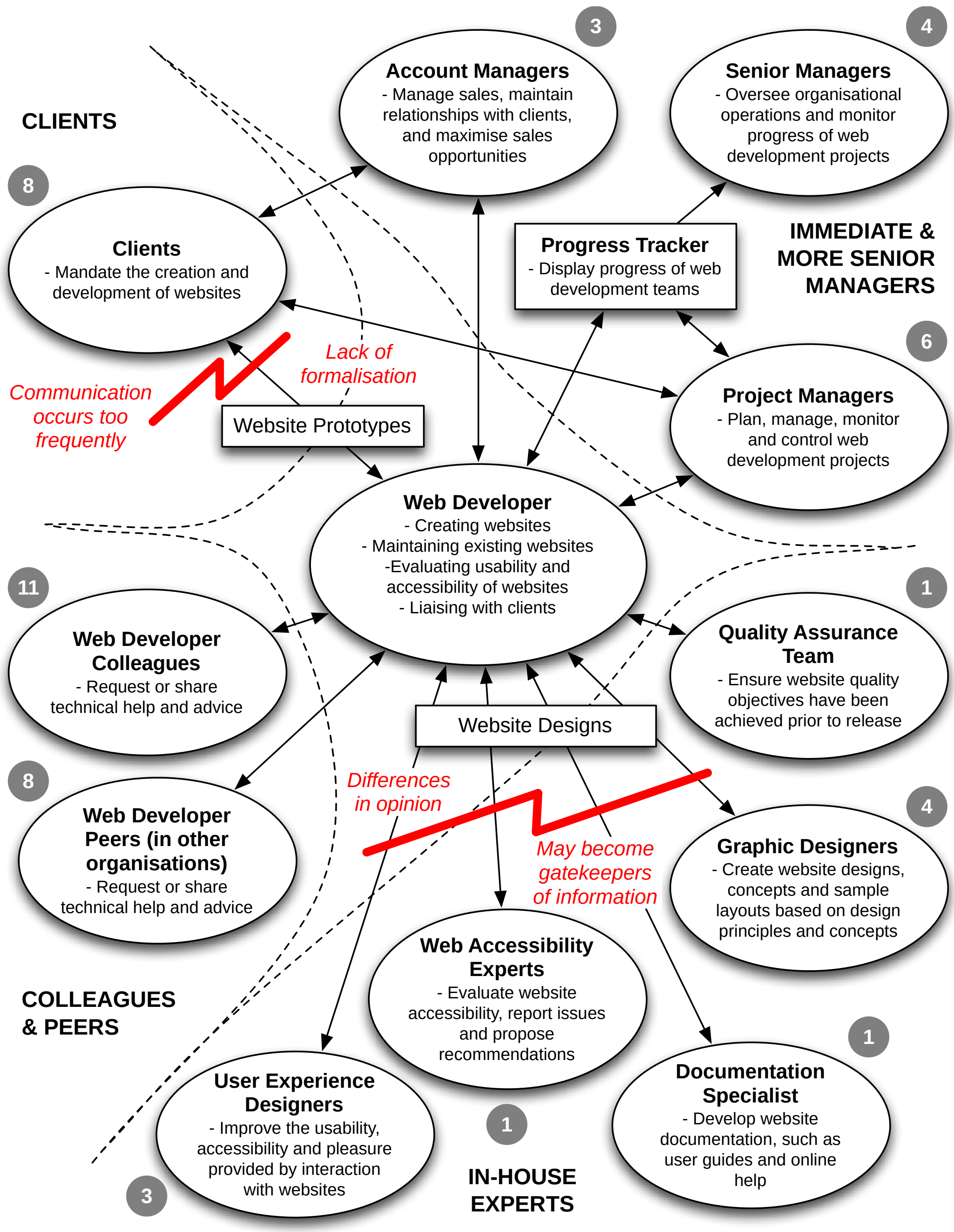


Figure 4.1: Flow Model

Bubbles represent people or groups and are annotated with lists of what is expected of the roles. Counts of the number of times a particular person or group was mentioned by participants in this study are displayed next to each bubble. There is some overlap between these counts, as participants typically mentioned more than one person or group. Arrows between bubbles represent the communication that occurs between people or groups to get work done. Small boxes on arrows represent artefacts. These may be physical objects, such as documents or messages, or conceptual, such as conversations, or announcements. Large boxes represent places that people go in and out of in order to get their work done. Problems in communication or coordination are represented as red lightning bolts.

4.3.1.2. Immediate and more senior managers

The majority of web developers in this study interact with immediate and more senior managers in their organisation. Six participants work closely with project managers, who are responsible for planning, managing, monitoring and controlling web development projects and liaising with clients or other departments in the organisation. Four participants interact with more senior managers, who oversee the organisation's operations. Three participants interact with account managers, who manage sales, maintain relationships with clients, and maximise sales opportunities.

Interaction with project managers occurs fairly frequently, especially during the lead-up to a website launch. Project managers are typically seated close to web developers, allowing interaction to occur face-to-face. Interaction with more senior managers occurs less frequently, usually during weekly or fortnightly face-to-face meetings. Web developers rarely ask more senior managers for technical advice but may turn to them for assistance with business problems. Interaction with account managers is intermittent and usually conducted via email.

Communication with both immediate and more senior managers often occurs indirectly, via online project management tools or physical progress charts. These are used in many organisations to track the progress of web development teams. This fairly hands-off approach to project management, in which web developers are trusted to communicate their progress and management rarely interferes, means that breakdowns rarely occur between web developers and either immediate or more senior managers.

4.3.1.3. Clients

The larger the organisation, the more distanced web developers tend to be from clients. Five participants in this study have no client interaction whatsoever, with project managers and account managers shouldering this responsibility. The remaining eight web developers interact with clients directly. This often involves gaining feedback on website prototypes or providing technical support and advice.

Client interaction occurs through a variety of communication channels, including email, phone calls and conference calls, IM, fora, and, occasionally, face-to-face meetings. Some methods of communication, such as email, can be particularly time-consuming or lack formality, which can lead to breakdowns. One participant complained about hearing from clients "every day, every five minutes!" [WD12]. Another participant said, "there are few days I don't talk with the client", which, he added, is "bad, because communicating is time-consuming!" [WD11]

Constant client communication is a particular concern for the freelance web developer in this study, who often struggles to distinguish his work life from his private life:

> With the Internet being 24/7, they think that you're 24/7 and suddenly you're taking calls when you're having tea with your family or out with friends. It may only take 5 minutes but you're constantly thinking, 'I hope my phone doesn't go off' [WD7].

Web developers who communicate directly with clients often face disruption, due to frequent demands on their time. This can lead to communication breakdowns and often results in web developers seeking less intrusive communication methods.

4.3.1.4. In-house experts

Many of the web developers in this study interact with in-house experts employed by their organisation. Four participants interact with a graphic designer; three participants interact with a UX designer (who are often responsible for web accessibility); one participant interacts with a QA team; one participant interacts with a documentation specialist; and one participant interacts with a dedicated accessibility expert. Surprisingly, none of the web developers in this study interact with external experts, perhaps due to the perceived costs of such services, particularly for SMEs. Interaction with in-house experts is almost always conducted via email and involves sharing and discussing website designs and prototypes.

Though in-house experts provide specialist support to web developers, they also have the potential to interrupt web developers' workflows. Breakdowns frequently occur between web developers and in-house experts over differences in opinion. Some web developers resent having to run even the smallest change past other people in the organisation.

The presence of in-house experts however, appears to result in some web developers absolving themselves of certain responsibilities. This is particularly the case for web accessibility, with two participants admitting they do not want the responsibility of making their websites accessible. One participant feels his proximity to a website means he is not in a position to determine its accessibility and that it often needs: "a fresh pair of eyes, someone who is completely new to look at it" [WD3]. Another participant does not feel in a position to make decisions about website accessibility and prefers to be given a very precise specification instead. He said: "I don't have to deal with the decision about where to show this piece of information or organise this menu. These are not things I care much about" [WD11].

Because some web developers are reluctant to take responsibility for web accessibility, in-house experts often become the only source of accessibility information in an organisation. One participant said "at the moment it's kind of up to [the accessibility expert] whether it's accessible or not and she generally gives her opinion on it" [WD3]. This gatekeeping of information can result in breakdown between web developers and in-house experts. One participant, who also manages a development team, said:

> Having one person as the arbiter of accessibility makes their position quite difficult because they're not necessarily able to explain why something has to be done a particular way, only that it has to be done a particular way, and that can create a personal tension that you don't necessarily want [WD4].

Though some web developers welcome the involvement of in-house experts, particularly with regards to creating accessible websites, breakdowns often occur over differences in opinion and the gatekeeping of information.

4.3.2. Cultural model

The cultural model (see Figure 4.2) represents the influences and pressures present in the work environment and the impact they have upon people's expectations, desires, values, and relationships.

In addition to identifying a self-imposed pressure by web developers to adhere to good standards of practice, I also identified three other broad "influencers" (that is, the people, organisations and groups who affect or constrain work): the web developer's organisation; the web developer's clients; and the web development industry and standards bodies.

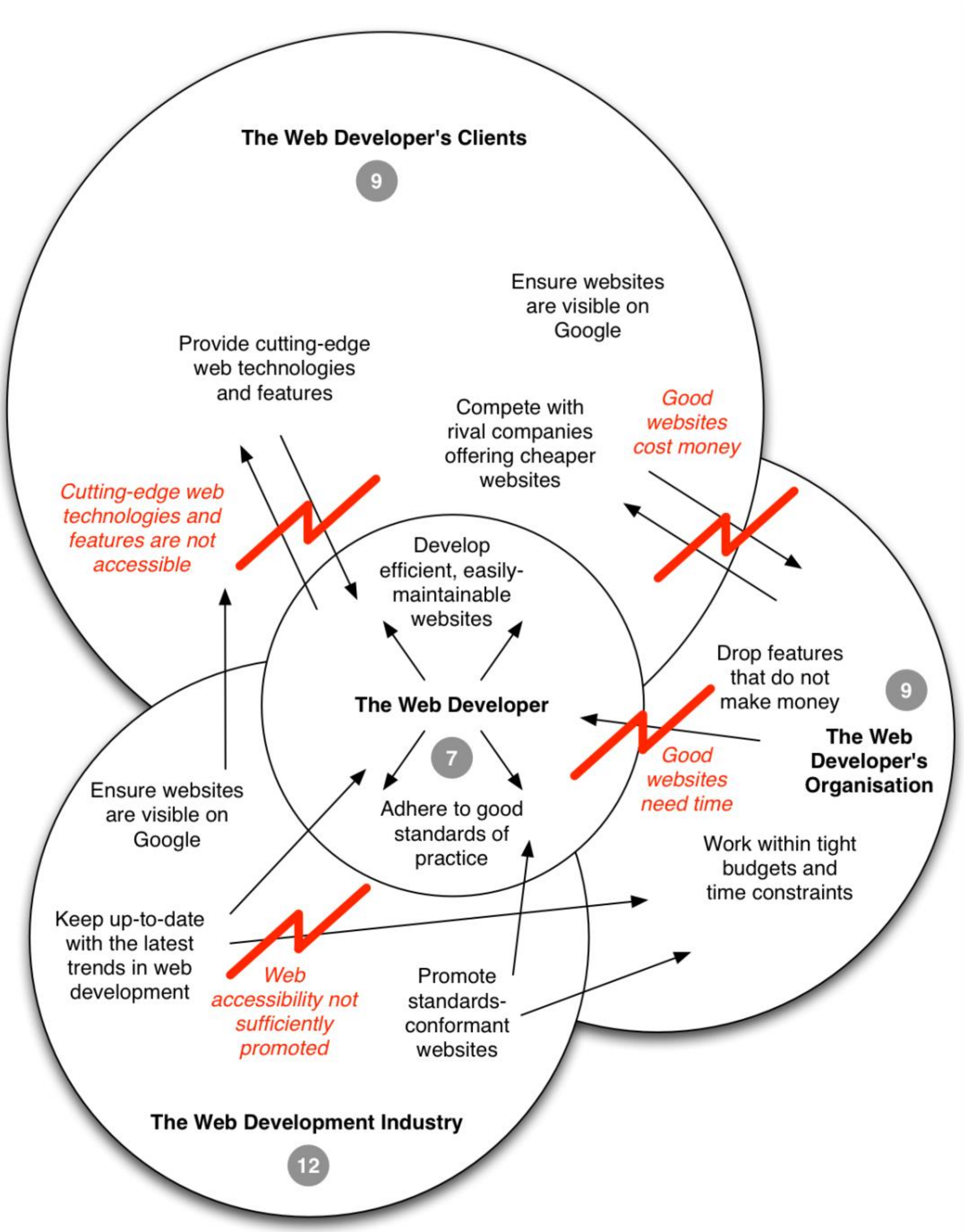


Figure 4.2: Cultural Model

Bubbles represent 'influencers': the people, organisations and groups who affect or constrain work. Counts of the number of times a particular influencer was mentioned by participants in this study are displayed next to each bubble. There is some overlap between these counts, as participants typically mentioned more than one influencer. Arrows represent the direction of influence and how pervasive it is. The extent of the effect on the work is shown by the amount the bubbles overlap. Particularly harmful problems, which severely interfere with the work, are represented as red lightning bolts.

4.3.2.1. The web developer

Seven web developers referred in some way to a self-imposed pressure to develop efficient, easily maintainable websites and adhere to good standards of practice, which often includes web accessibility.

Five participants described being motivated by a particular sense of responsibility towards people with disabilities. One participant said, "I will try to make good code as best I can. It is important to have a website that can be used by all" $_{[WD12]}$. Another participant said: "most things that aren't accessible that could have been accessible will be the developer's fault" but acknowledged that other stakeholder groups are involved, adding "everyone has got to dance in time" $_{[WD2]}$.

Some web developers are motivated to create accessible websites even when clients have not explicitly requested it. One participant described how it is possible to "smuggle" in accessibility by "saying it is good coding practice, good for maintainability, finding bugs, saving time" $_{[WD1]}$. Another participant admitted: "sometimes it's not even for those who can't see. Sometimes it's nice to have that minimal code for things like mobile phones. The page loads much quicker" $_{[WD7]}$.

Ensuring that websites are accessible to everyone, but in particular to people with disabilities, appears to provide considerable motivation to many web developers and few breakdowns occur as a result of this self-imposed pressure.

4.3.2.2. The web developer's organisation

Nine web developers referred in some way to the influence that their organisation has upon their work. This influence stems from individuals and groups within an organisation, such as immediate and more senior management, in-house experts, colleagues, as well as the overall organisational culture.

Smaller organisations appear to promote a more egalitarian culture. One participant, who works in an SME, described how "everyone has an equal say and everyone's opinions count" $_{[WD9]}$. The culture in larger organisations appears to be more constrained. One participant, who works in a large enterprise, described how "I can express my views and tell [project managers] how to do things, but ultimately I leave the decisions up to them if they insist, because it's their project" $_{[WD1]}$.

Irrespective of the size of their organisation, the majority of web developers in this study appear to be under considerable pressure to work within tight budgets and time constraints. Breakdowns occur when web developers are forced to scale back the complexity of websites and drop less discernible services, such as SEO, standards conformance, and web accessibility. One participant revealed:

> If we have to deliver something quickly, we just try to do the simplest thing possible and drop a number of details that you don't have time to deal with … and one thing could be accessibility. You do something that is easy for you and close to what you think the customer wants to see $_{[WD11]}$.

Web accessibility is not promoted at all in the majority of participants' organisations and is only considered if a client requests it, which they seldom do. One participant, working in higher education, finds the lack of awareness from his organisation particularly frustrating and often struggles to ring-fence the necessary time and resources to make websites accessible. Another participant said, "It's all about time and just the fact that it's not in the requirements. I'd actually be doing the business a disservice if I spent too much time on it" $_{[WD1]}$.

Developing good websites needs time and costs money. Although some organisations promote a more egalitarian culture the others, the considerable time and budgetary pressures that most organisations place upon web developers often means certain 'invisible' features, including web accessibility, are dropped.

4.3.2.3. The web developer's clients

Nine web developers referred in some way to the influence of clients upon their work.

Satisfying client demands and managing their expectations can be a major source of frustration for web developers and a common cause of breakdowns. Clients place a considerable amount of pressure on web developers to ensure their websites are visible on Google and include cutting edge web technologies and features. This is often at the expense of making the websites accessible. One participant described how: "The more complicated, the more widgets and interactive things, the more impressive the page gets, the more the accessibility suffers" $_{[WD12]}$.

Breakdowns also occur when web developers who strive to make their websites accessible are undercut by rival web developers, who are willing to sacrifice accessibility in order to offer a cheaper website. According to one participant, "All the really clever stuff is invisible, so it is difficult to charge for. I know I'm doing it but does the client know I'm doing it? No." [WD7]. Another participant said:

> People will turn around and go 'oh, I can get this website knocked up by such and such down the road, and he'll do it in three days, and it'll only cost me this much money, so, y'know, can't you do that?' Yes, we can, if we don't bother with all the bells and whistles, which you probably won't notice [WD2].

Although clients often impose considerable pressure upon web developers to ensure their websites include the latest cutting-edge features, this rarely includes making websites accessible.

4.3.2.4. The web development industry and standards bodies

All except one of the web developers in this study referred in some way to the influence that the web development industry and standards bodies have upon their work. This influence stems largely from the other web developers in the industry, who share and discuss latest web technologies and techniques, and standards bodies, which define current web standards and specifications. It also stems from browser manufacturers, who provide platforms for the latest web technologies; and the technology company, Google, which has a powerful influence on the selection of websites presented to users.

Web developers face a constant pressure to keep up-to-date with the latest trends in web development and further their knowledge of the field. Little formal training is provided by organisations, often due to the considerable cost. This means that the majority of web developers in this study are self-taught and rely upon online tutorials to further their knowledge. Each participant in this study feels his knowledge of web development is either up-to-date or only slightly "behind the curve". While they frequently described web development as a "constantly changing field", many noted that this is a good thing as it keeps them "on their toes".

Efforts by standards bodies to promote a culture of good practice (such as accessibility statements and conformance "badges" that allow web developers to declare conformance to various web standards) receive a mixed reception. Some participants feel they provide an indication of their commitment to web standards and are useful for advertising an organisation's approach to web accessibility. Others feel they are dated and, because they rely upon self-certification, are often used falsely or incorrectly. One participant said, “If you have an accessibility problem and you want the site to work, the fact that there is an accessibility statement wouldn’t change whether the site works for you or not” [WD6].

According to one participant, the promotion of web accessibility in the web development industry could be improved, particularly at industry conferences, to ensure that web developers are more aware of people with disabilities. The participant recalled watching a presentation at a web development conference by the technology expert, Robin Christopherson, who himself is blind:

> He demonstrated what it is like for him to browse the web, which was really quite shocking. He called out Facebook as being absolutely terrible with a screen reader. That was the big realisation point. Actually seeing how blind people use [the web] was the big eye-opener and raised the importance of accessibility [WD6].

Despite imposing considerable pressure on web developers to keep up-to-date with the latest web technologies and declare conformance to web standards, it appears the web development industry and standards bodies could do more to promote web accessibility to web developers.

4.3.3. Physical model

The physical model (see Figure 4.3) represents the physical environment in which the work takes place, incorporating the desks and workstations, rooms, floors, and buildings in which people move about and work.

I discovered that the physical environment appears to have little bearing upon the web developers in this study. This is because the bulk of their work takes place at a computer. Web developers use few physical artefacts and, aside from a computer, their desks and workstations remain fairly sparse. Five web developers keep printed project documents on their desks for reference and some pin up so-called “cheat sheets” (short documents packed with useful commands and information, such as common HTML elements or CSS attributes). The majority of web developers however, keep files, documents, and other work-related artefacts on their virtual, as opposed to physical, desktops.

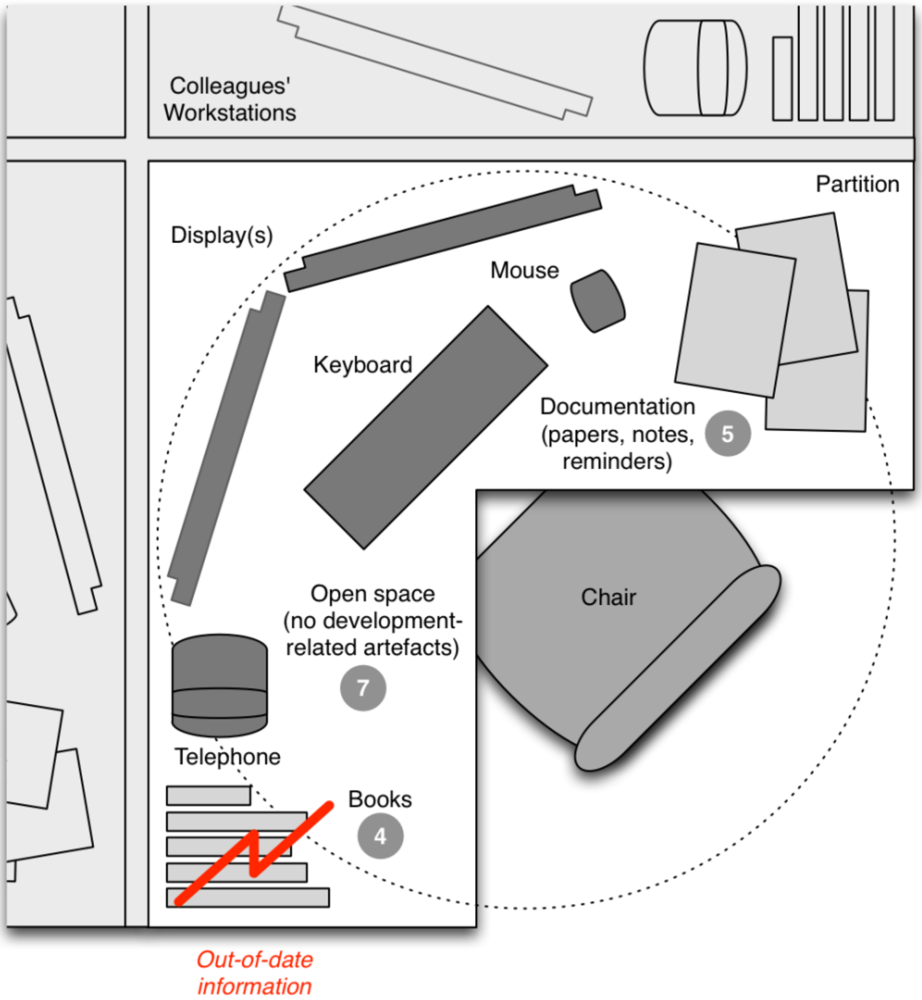


Figure 4.3: Physical Model

This represents the (limited) physical environment in which web developers' work takes place. It includes hardware, software and other artefacts that are present in the space and displays how they are laid out in relationship to each other. Counts of the number of times a particular artefact was mentioned by participants in this study are displayed next to them. There is some overlap between these counts, as participants typically mentioned more than one artefact. Problems that show how the physical environment interferes with the work are represented as red lightning bolts.

Four web developers keep a small collection of technical reference books on their desk but admit that they rarely consult them. Rather than opening books containing potentially dated information, they typically turn to Google for up-to-date information. One participant said, "there's a book there, which I've probably never looked at … It's all on the computer" [WD2]. One participant however, revealed, "I'm an old schooler, I have my programming books back from when people bought books. They're always the best because they get down to the architecture, not the fluff at the top" [WD11].

Although many organisations use online project management tools to monitor the progress of web development teams, some still use physical whiteboards or magnetic displays. Gantt charts illustrate the start and finish dates of different project components and 'burn down charts' represent the amount of work left to do versus time. Physical progress trackers such as these are typically located close to where web development teams sit, and web developers manually update them as they achieve their tasks. Though some web developers find it a hassle to physically update the progress tracker, this is often countered by the satisfaction of viewing progress.

With the bulk of web developers' work taking place online, few breakdowns occur as a result of their physical environment.

4.3.4. Sequence model

The sequence model (see Figure 4.4) represents the sequence of activities that form work tasks, the overall intent and subintents behind each sequence, and the triggers that initiate them.

During the contextual inquiries, I prompted participants to carry out certain tasks related to the development of websites, including previewing webpages, validating webpages, and evaluating the accessibility of webpages. Of particular interest to this investigation was the sequence of activities involved in evaluating the accessibility of webpages. The overall intent to evaluate the accessibility of a website is broken down into several subintents: identify webpages to evaluate; conduct automated accessibility evaluation; conduct manual accessibility inspection; and refer website to accessibility expert.

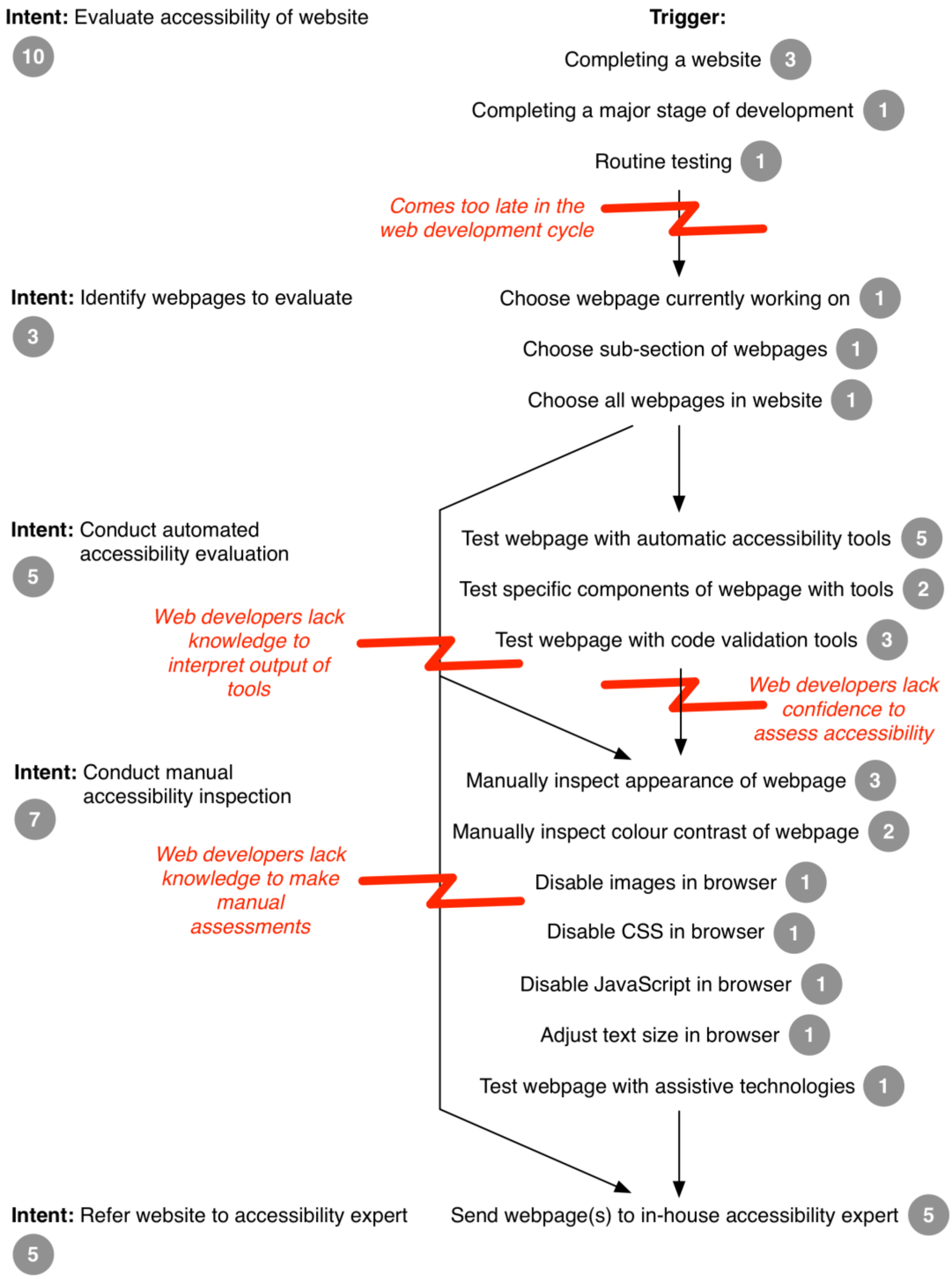


Figure 4.4: Sequence Model (of evaluating the accessibility of webpages)

This begins with the overall intent of the sequence and the trigger that initiates it before listing each step in order. Counts of the number of times a particular step was mentioned by participants in this study are displayed next to them. Loops and branches reflect patterns of work. Any steps that cause problems are labelled with red lightning bolts.

4.3.4.1. Evaluate accessibility of website

Although ten web developers in this study demonstrated how they might evaluate the accessibility of webpages, few had an established procedure.

Web developers tend to only consider accessibility testing towards the end of the development process. For three participants, accessibility testing is usually triggered by the completion of a website. One participant attempts to conduct routine accessibility testing on each of the (completed) websites for which he is responsible. Another participant attempts to conduct accessibility testing after each major stage of development, followed by an overall check at the end of a project. No other participants specified a trigger for accessibility testing.

Breakdowns occur over the lack of prominence of web accessibility in the development process. Six participants would prefer it to be taken into consideration throughout the development process. This is because retrospectively tackling web accessibility problems at the end of the development process can require a considerable amount of effort. One participant said that web accessibility is: "The kind of thing you need to know about beforehand really because it's a pain to change it afterwards" [WD2].

Embedding web accessibility through the development process would, according to one participant, embody the entire development team with the appropriate ethos. He said "I think everybody has to understand or at least take account of accessibility at all of the levels [of web development]" [WD4]. He went on to describe how web accessibility "has traditionally been seen as 'oh and now let's do the accessibility work' … and I think embedding accessibility in the whole process would then mean that it wouldn't be overlooked, it wouldn't be the last thing to look at" [WD4].

Leaving web accessibility until the very end of the development process results in a considerable amount of extra work for web developers. Integrating web accessibility throughout the development process would make the task more achievable.

4.3.4.2. Identify webpages to evaluate

Only three participants described how they identify a sample of webpages to test. One participant chooses to test whatever webpage he is currently working on. Another chooses a sub-section of webpages to test (but provided no further details). Still another claimed to test all webpages within a particular website.

4.3.4.3. Conduct automated accessibility testing

Having established the need for accessibility testing, and identified which webpages to test, five participants use automated web accessibility evaluation tools, such as Total Validator[40], AChecker[41] or Testo Accesibilidad Web (TAW)[42] (see Section 2.1.1 and 3.3.5). One participant claimed to regularly use an automated evaluation tool but when asked to demonstrate it, he discovered it had actually been discontinued for several years and no longer worked.

Three participants only run their websites through a basic code validation tool, such as the W3C Validation Service[43], to check whether they conform to the HTML standard for accessibility testing. One said "to be honest, after you've got validated HTML, you can be reasonably confident that it's not atrociously inaccessible" $_{\text{[WD3]}}$. While valid code certainly contributes to the accessibility of a website, this comment suggests a limited understanding of the subject.

Two participants use tools that automatically test a specific component of accessibility. For example, one participant uses an online tool that alters the colours on a webpage to simulate colour blindness. The same developer also uses a browser extension called Fangs[44], which emulates the screen reader, JAWS[45]. Without an understanding of how such tools relate to actual users however, web developers may place too much confidence in the feedback they provide. For instance, one participant uses an online tool to determine whether the contrast between foreground and background colours is sufficient. He said, "That's the only sort of accessibility testing I have done. I would be relying on this [tool] to say 'yes, this is fine' or 'no, this not'" $_{\text{[WD6]}}$.

The use of automated accessibility evaluation tools appears to contribute little to web developers' understanding of web accessibility and may actually promote a false sense of

---

40 https://www.totalvalidator.com
41 https://achecker.ca
42 http://www.tawdis.net
43 https://validator.w3.org
44 http://www.standards-schmandards.com/projects/fangs
45 http://www.freedomscientific.com/Products/Blindness/JAWS

confidence as web developers lack the expertise to accurately interpret the output of such tools.

4.3.4.4. Conduct manual accessibility inspection

Seven participants use a haphazard collection of manual checks. For instance, five participants rely upon a basic knowledge of web accessibility and 'rules of thumb', such as checking images for alternative descriptions or choosing appropriate background and foreground colours. One participant said, "If I like it, I hope the end user will like it" $_{[WD13]}$. Another participant carries out a "quick visual inspection" claiming that "some things you can do visually: look at it, see that you have good colour contrast, flick around with text size" $_{[WD1]}$. One participant does not carry out any formal accessibility testing, "other than the obvious stuff just by looking at it". He relies upon "general rules that I've picked up from places and remembered from other projects", adding, "I think, partially, it's just common sense" $_{[WD3]}$. These comments demonstrate that the web developers do not actually understand the detailed process necessary to conduct manual accessibility inspection.

Three participants use browser extensions, such as the Web Developer toolbar[46], to disable different components of webpages. For example, one participant disables JavaScript but "probably not much beyond that" as "you rely more on the fact that you built those things in as you've gone along rather than test them afterwards, because quite often there's nothing you can actually test" $_{[WD2]}$. Another participant feels that disabling images and CSS in the browser is a sufficient test of accessibility. He said, "You develop an innate instinct for what is going to work and what isn't" $_{[WD7]}$. Again, these comments indicate that the web developers rely more on intuition than knowledge when it comes to web accessibility.

Only one participant uses ATs to check the accessibility of his websites. He attempts to test the accessibility of websites using JAWS screenreader[47], ZoomText screen magnifier[48], and Supernova screenreader/magnifier[49]. Consequently, he feels "fairly confident that what I chuck out there is accessible". The participant did however

[46] https://chrispederick.com/work/web-developer
[47] http://www.freedomscientific.com/Products/Blindness/JAWS
[48] https://www.aisquared.com/products/zoomtext
[49] https://yourdolphin.com/supernova-magnifier-screen-reader

acknowledge the considerable cost of these applications and admitted that it was: "not a typical solution" [WD5].

Though web developers use a variety of manual accessibility tests, their lack of web accessibility expertise means the tests are little more than guesswork and speculation.

4.3.4.5. Refer website to accessibility experts

Five participants do not conduct any accessibility testing themselves but instead refer their websites to an in-house expert, such as a UX designer or, in one organisation, a dedicated accessibility expert.

This reliance upon in-house experts appears to be caused by a lack of confidence on the part of web developers. Three participants admitted they do not feel capable of making any assessments of accessibility, particularly for issues that cannot be 'programmatically determined' by web accessibility evaluation tools and that require subjective interpretation. One participant said, "if it is an accessibility issue that can't be tested automatically, I'm 99% sure that I won't be able to judge it." Referring to a "developer mindset", he said:

> I am not the best person to make a website accessible because I often realise that when I take the decision on my own about how to organise the page or the functionality or the menu and then I show it to a user… 90% of the time the user says 'it would be better if you did it like this'. I look at what is easier for me, so I do things that way, not realising it makes it less accessible than another way. So developers don't have the skills for assessing accessibility [WD11].

The majority of participants consider the main problem in the creation and evaluation of accessible websites to be a lack of awareness from web developers. One participant said, "if accessibility is being overlooked, I'd like to think it is not due to the laziness of the developer, I'd like to think it is because they are just not aware of it" [WD6]. The same participant attributes this lack of awareness to "a lack of good information that is easy to find, and information that promotes awareness". He feels that "accessibility is not something that is particularly well documented. I don't feel like there is an accessibility resource" [WD6].

The tendency for web developers to refer their websites to dedicated accessibility experts appears to stem not from a shirking of responsibility but from a lack of knowledge and confidence in their ability to accurately assess the accessibility of websites.

4.3.5. Artefact model

The artefact model (see Figure 4.5 and Figure 4.6) represents the (typically physical) objects that people create, use and modify in the course of doing work.

With the exception of physical whiteboards or magnetic displays to monitor progress, web developers rarely use physical artefacts, as the bulk of their work takes place online. They do, however, use certain "virtual" artefacts. Of particular interest to this investigation were the artefacts relating to web accessibility. These include guidelines and specification documents, such as WCAG 2.0 (Caldwell et al., 2008a), as well as the output and feedback from web accessibility evaluation tools, such as Total Validator, AChecker or TAW. I identified three distinct ways in which these artefacts affect or constrain the work of web developers: the volume and comprehensiveness of accessibility information; the organisation and structure of accessibility information; and the language and terminology of accessibility information.

4.3.5.1. Volume and comprehensiveness of information

Six participants referred in some way to the volume and comprehensiveness of information presented by existing accessibility tools, guidelines and resources. Though web developers acknowledge the substantial amount of work involved in making websites accessible, they find the amount of information overwhelming and the number of items to test unrealistic.

Breakdowns commonly occur over the verbosity of the output from web accessibility evaluation tools. In addition to presenting the pass-fail outcome of tests that can be evaluated automatically, web accessibility evaluation tools often generate a large number of optional "warnings", highlighting aspects of webpages that must be checked manually. This combination of automatic and manual tests often results in a considerable amount of feedback, which some web developers find overly verbose and time-consuming to process. One participant considers automatic accessibility evaluation tools to be "so over-zealous that you can't see the wood for the trees" $_{[WD2]}$. The participant adds: "Sometimes you get onto the accessibility checks and there's a grey area in there and you end up with pages and pages and pages of noise" $_{[WD2]}$.

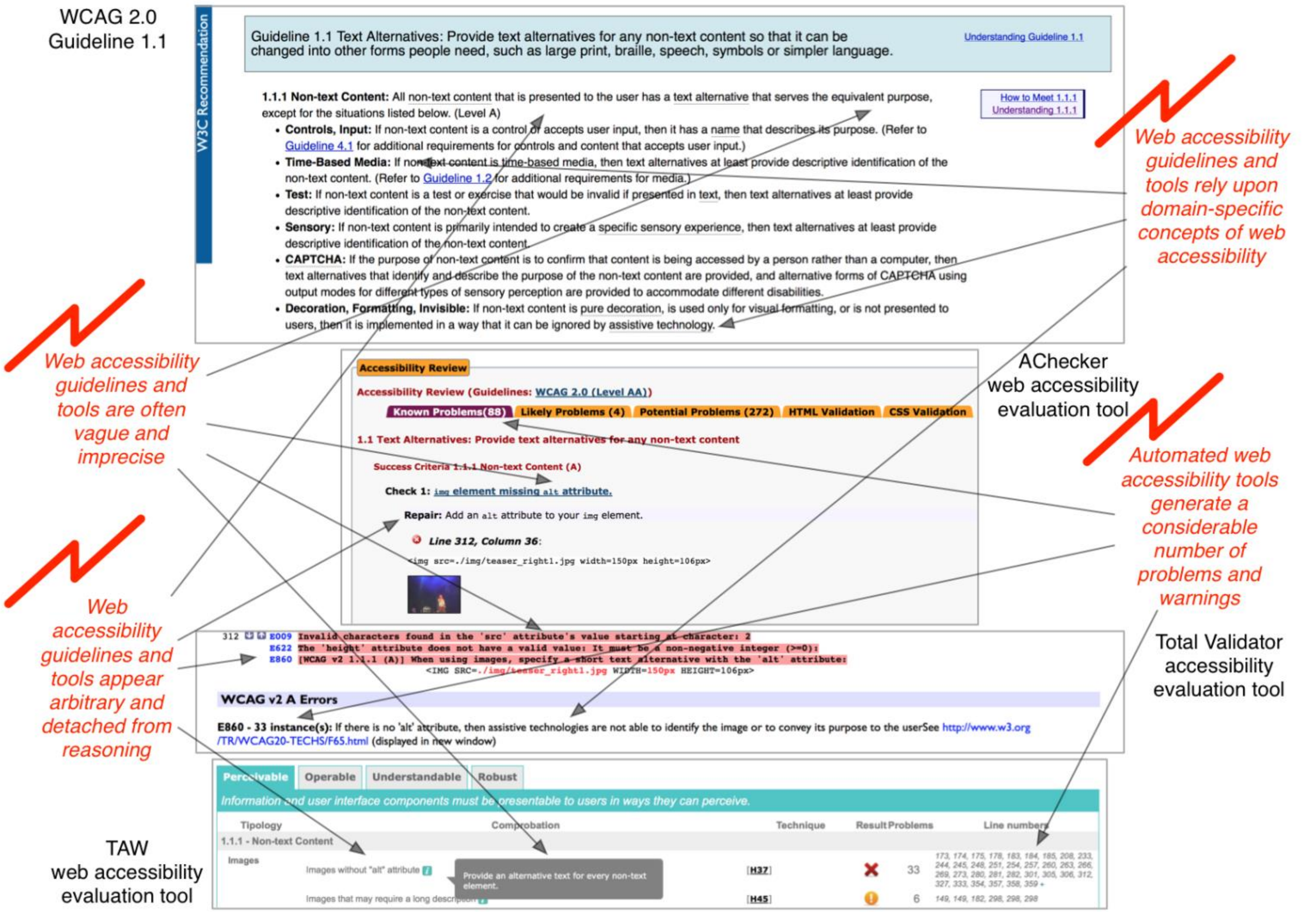


Figure 4.5: Artefact model (1 of 2)

The artefacts relevant to this study include guidelines and specification documents, such as WCAG and the feedback from automated web accessibility evaluation tools. Problems in using the artefact are represented as red lightning bolts.

Screenshots of WCAG 2.0, Total Validator, and TAW reprinted with permission. Screenshot of AChecker reprinted under the GNU Lesser Public Licence, version 2.1.

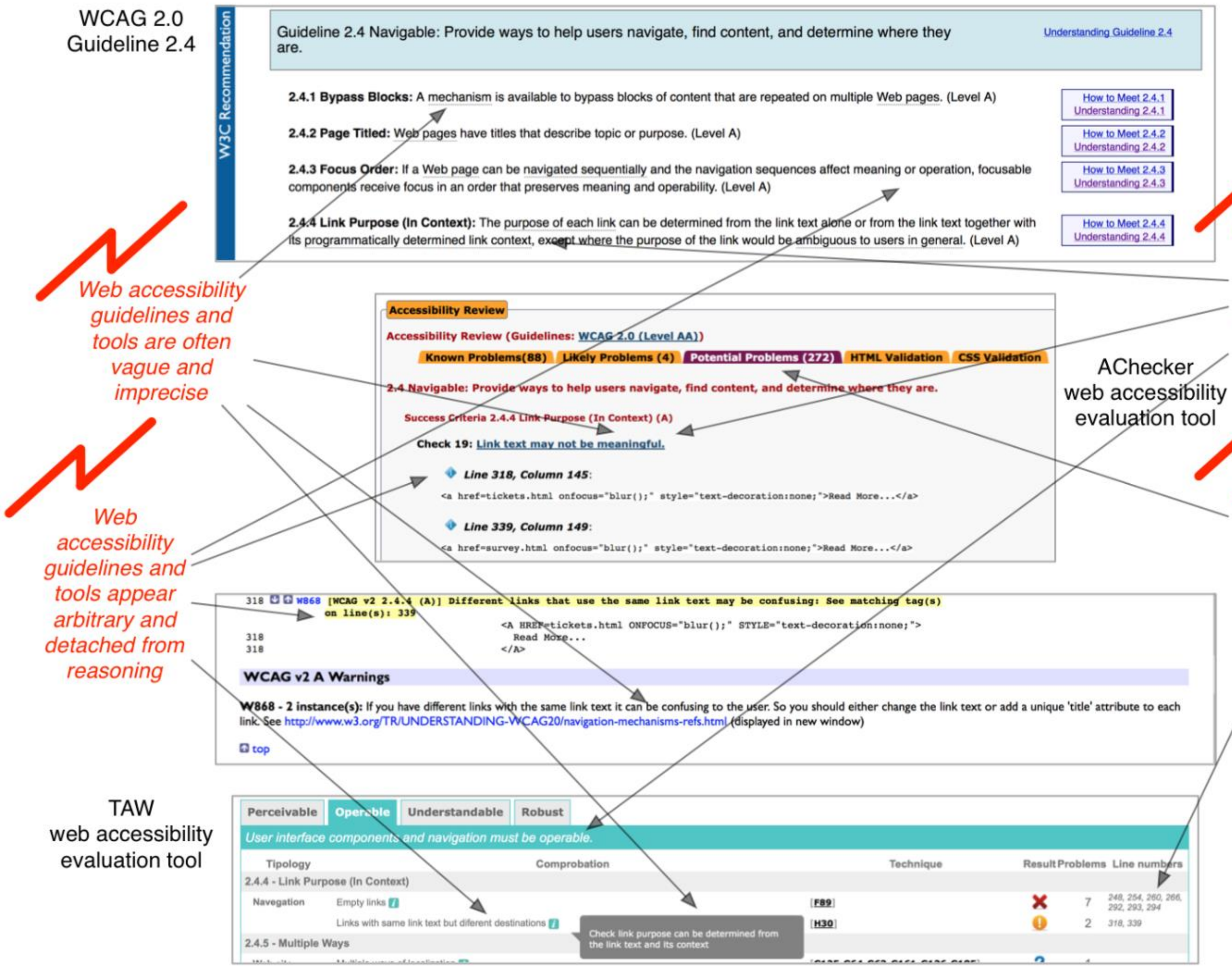


Figure 4.6: Artefact model (2 of 2)

The artefacts relevant to this study include guidelines and specification documents, such as WCAG and the feedback from automated web accessibility evaluation tools. Problems in using the artefact are represented as red lightning bolts.

Screenshots of WCAG 2.0, Total Validator, and TAW reprinted with permission. Screenshot of AChecker reprinted under the GNU Lesser Public Licence, version 2.1.

4.3.5.2. Organisation and structure of information

Nine participants referred in some way to the organisation and structure of existing web accessibility tools, guidelines and resources, which tends to be different to how web developers approach the creation of websites. While web developers' practices are typically related to the code they are working on, the guidance and information that accessibility tools, guidelines and resources deliver tends to be organised according to abstract, accessibility-specific groupings (such as WCAG 2.0's POUR principles). Web developers find this way of organising the information difficult to navigate and apply to their work.

Web developers would instead prefer the organisation of accessibility guidance and information to be more closely related to the code they are working on. This would improve the navigability of tools, guidelines and resources and make it easier for web developers to apply the guidance and information to their work. For many participants, this would be in the form of a checklist, similar to the output of code validation tools, such as the W3C Validation Service. One participant said, "Ideally, I'd love some kind of really quantitative list of all the things that are wrong. I'd want to go through it line-by-line saying, 'right, this is wrong, okay fix it, revalidate'" [WD1]. Another participant would prefer "something I can work off, something I can check, in black-and-white, whether something ticks boxes or not" [WD3]. According to another participant: "a checklist would be quite useful" because "it doesn't feel like there is a list anywhere on the web or a complete guide to accessibility" [WD6].

Six participants also find that existing tools, guidelines and resources provide very little information about the people who benefit from accessible websites. One participant stated: "There seem to be a lot of what we call 'diktats', which are effectively recommendations, but without any real explanation" [WD4]. Another participant described how "some guidelines seem more spurious or for the sake of it than others". He added "there are things I get told to change by Total Validator that I will happily do but haven't got a clue why" [WD2]. The consequence of this is that the information appears abstract, arbitrary and unrelated either to the web developers' work or to the people it is intended to benefit. Providing the reasoning behind accessibility guidance would encourage and motivate many web developers. One participant explained why this justification is necessary:

> If you actually gave a developer a reason, I think, more than anything, that would probably make them more aware and realise that by not doing this you are omitting every blind person out there from using your website. Something along those lines would give you a very valid reason for doing it as opposed to just trying to hit a standard. [WD8].

4.3.5.3. Language and terminology of information

Eight participants referred in some way to existing web accessibility tools, guidelines and resources not speaking their language. Instead, tools, guidelines and resources tend to rely upon vague and indefinite statements that assume web developers are familiar with accessibility-specific concepts. These statements include undefined directives, such as 'provide users enough time to read and use content' (WCAG 2.0, Guideline 2.2) as well as optional warnings, such as 'Link text may not be meaningful' (aChecker, potential problem 19), both of which web developers find unhelpful, ambiguous and off-putting.

Web developers would instead prefer to receive clear, precise and unambiguous accessibility guidance and information. One participant feels such information should be "as specific as possible as to the exact nature of the problem, the location of the problem, and the best way to address the problem" [WD2]. He added: "Rather than something saying 'ooh, you've used an image tag and it's not accessible', it should say 'to make it accessible go and put some text describing the image in the alt tag.' Pretty black and white instructions, basically!" [WD2]. Another participant explained:

> If you just give me 'you have to break paragraphs to make them visible at that screen size', well, okay, that's a requirement I can deal with. But if you give me a more detailed requirement it is easier. With the previous one, I have to guess a number of things and may get it wrong. Whereas if you give me a more specific requirement, that's fine [WD11].

4.4. Discussion

This chapter has presented a contextual inquiry investigation into the working practices of 13 professional web developers. The aim of this study was to examine the role of professional web developers and their working practices in much greater detail to understand why they are struggling to make websites accessible and identify how best to support them. The contextual design methodology provided a comprehensive means of exploring the complex working practices of this cohort.

Consistent with previous surveys of web developers (e.g. Petrie, 2006; Korn, 2009; Farrelly, 2011; Yesilada et al., 2015), the difficulties that the web developers in this study face in creating accessible website are rarely due to a lack of awareness or concern. Web developers' awareness of disabled people and their motivation to support them provides an important influence upon the accessibility of the websites they develop. Many web developers consider web accessibility to be a particular example of best practice and are driven by the desire to ensure their websites are accessible to as many people as possible.

Some web developers are particularly keen to learn about the beneficiaries of web accessibility and understand how best to support them. They feel it is important to be aware of the reasoning behind any accessibility guidance, as opposed to indiscriminately following arbitrary and unjustified recommendations. These findings present a much more positive outlook than the studies of web developers by Vora (1998), Rosson et al. (2005), and Freire et al. (2008), in which few respondents had ever considered the accessibility of their websites, let alone thought about the people using them.

Furthermore, in the majority of organisations visited in this study, web developers are, or at least, believe themselves to be, the stakeholder group most responsible for creating and maintaining accessible websites. This supports the findings of Lazar et al. (2004) in which 82% of web developer respondents considered themselves responsible for web accessibility. Nevertheless, web developers in both the Lazar et al. study and the current study stressed that web accessibility is not an individual effort and that it requires input and cooperation from a range of stakeholder groups.

Despite their good intentions and positive attitudes regarding web accessibility, the web developers in this study appear to have little theoretical understanding of disability issues or practical knowledge of how best to provide accessibility support. Their responses suggest that inaccessible websites are rarely created out of negligence or malice, but instead out of unfamiliarity with the ways in which disabled people use the web and the problems they encounter. For example, one web developer admitted having little understanding of the rationale behind established accessibility recommendations, while another referred to the recommendations as "diktats" that are imposed without justification. Another web developer described how he had only really grasped the importance of web accessibility following a conference presentation by a blind speaker who had demonstrated the challenges faced by people with visual impairments.

Without a solid theoretical understanding of disability issues, web developers haphazardly approach the practical aspects of web accessibility evaluation. For example, several web developers admitted relying on intuition and half-remembered rules of thumb to determine the accessibility of websites, while one considered it to be simply "common sense". Similarly, automated evaluation tools are routinely obeyed by many of the web developers in this study but with seemingly little understanding. Such findings are consistent with those of studies by Vora (1998), Rosson et al. (2005), Knight (2003), and the BenToWeb project (Petrie, 2006), which determined that web developers have little awareness of web accessibility problems, let alone the solutions available.

There are many well-established reasons for this unfamiliarity with web accessibility. For example, many of the web developers in this study are unaware of how to access training in web accessibility and lack the time to undertake it. Some of the participants feel that web accessibility support and information is neither sufficiently documented nor widely available and the majority of participants are unaware of existing web accessibility initiatives. These findings reaffirm those of studies by Knight (2003), Lazar et al. (2004), the DRC (2004), the BenToWeb project (Petrie, 2006), Freire et al. (2008), Korn (2009), Farrelly (2011), Hong et al. (2015) and Yesilada et al. (2015) in which many web developers attributed their own lack of expertise or knowledge of web accessibility to insufficient education and training materials.

The insufficient promotion and enforcement of web accessibility also results in web developers struggling to create accessible websites. Several web developers in this study feel that web accessibility is poorly promoted, either by individual organisations or by the web development industry as a whole. Although many web developers feel a moral and personal obligation to create accessible websites, there appears to be little legal or economic compulsion for them to do so. These findings support those of studies by Knight (2003), Lazar et al. (2004), Freire et al. (2008), Olalere and Lazar (2011), and Hong et al. (2015) in which ineffective legal enforcement and insufficient promotion of web accessibility are identified as considerable barriers to the development of accessible websites.

The attitudes and interests of clients is also, perhaps understandably, a dominant influence upon web developers' ability to create accessibility websites. This study reveals the frustration that web developers feel towards clients, who are often willing to sacrifice web accessibility in favour of cheaper and superficially similar websites. Though clients are usually eager for websites to include cutting-edge technologies, they are often reluctant to extend their eagerness to making websites accessible. These findings echo those of other studies by Vora (1998), Knight (2003), Lazar et al. (2004) the BenToWeb project (Petrie, 2006), Freire, Russo and Fortes (2008), Putnam et al. (2012), and Hong et al. (2015), all of which stress the considerable influence of client attitudes upon the accessibility, or inaccessibility, of websites.

One of the biggest influences upon web developers' ability to create accessibility websites, however, appears to be a lack of adequate knowledge and practical guidance on how to make websites accessible. Existing accessibility tools, guidelines and resources tend to overload web developers with an overwhelming amount of information that they find verbose and time-consuming to process. The information provided by existing tools guidelines and resources tends to be organised according to domain-specific groupings that web developers to difficult to navigate and apply to their work. The domain-specific web accessibility jargon used in tools, guidelines and resources also tends to be couched in vague and indefinite terms that web developers find unhelpful and off-putting. Furthermore, without any justification or rationale behind the information, web developers struggle to relate it either to their work or to the people it is intended to benefit. Existing tools, guidelines and resources are letting web developers down by not providing them with the support and information they need.

Without access to an appropriate amount of clear, concise and precise accessibility information that they can easily interrogate, understand, and apply to their work, web developers struggle to determine the accessibility of their websites. This manifests in low confidence and self-efficacy with regards to web accessibility. For instance, one web developer described how he "hope[d] the end user will like [his websites]" but appeared uncertain, or indifferent, about whether this is actually the case. Another admitted that he had little knowledge of web accessibility and that he relied solely upon automated tools to determine whether a website is accessible or not. Similarly, one web developer felt "99% sure" that if an accessibility issue could not be programmatically determined, he would be unable to judge it. It is evident that many of the participants in this study feel incapable of either making assessments of web accessibility or identifying appropriate solutions.

Web developers' low confidence and self-efficacy, in turn, results in an over-reliance on automated accessibility evaluation tools and web accessibility experts. While tools provide some degree of automation to web developers' workflows and absolve some of the responsibility for making assessments of accessibility, they contribute little to web developers' understanding of web accessibility and promote a false sense of confidence in some web developers. This finding supports the investigation by Petrie et al. (2007) into the usability of accessibility evaluation tools, which concluded that the tools do little to enhance web developers' understanding of accessibility issues. Similarly, previous studies of web developers (such as Lazar et al., 2004; Rosson et al., 2005; Lazar et al., 2003; and Trewin et al., 2010), found that web developers have little confidence in the feedback from tools. Accessibility experts also remove some of the responsibility for making assessments of accessibility. Their involvement, however, can result in both procedural and knowledge 'bottlenecks' in which they become gatekeepers of knowledge from whom web developers must seek approval. Ultimately, inadequate support and information results in web developers failing to integrate accessible web development practices into their existing workflows.

Existing web accessibility tools, guidelines and resources fail to deliver the necessary information to raise web developers' awareness of disabled people, further their knowledge of the techniques necessary to support disabled people, or develop their confidence about creating and evaluating accessible websites. Instead of time-consuming web accessibility tools that provide a false sense of confidence, or accessibility experts who remove the responsibility for assessing the accessibility of websites, web developers need access to information that gives them the knowledge, confidence and self-efficacy to make such assessments themselves. Whether such information is delivered via web accessibility guidelines or the feedback from web accessibility evaluation tools, it needs to educate and inform web developers without consuming too much of their time. Vague and indefinite terms must be replaced by precise, concrete and testable statements that are easier for web developers to apply. Similarly, domain-specific web accessibility jargon and terminology should be eschewed in favour of information that speaks the language of web developers. Crucially, the information must make a connection between the technical guidance it provides and the disabled people to whom it benefits, allowing web developers to better understand what they are doing and make confident and accurate assessments of accessibility.

The relatively intrusive nature of the contextual design methodology used in this study made participant recruitment particularly difficult. I approached a number of web developers who were initially keen to participate but, upon learning about what would be involved, become unwilling to allow me to conduct a contextual inquiry in their workplace. Other web developers who were willing to take part often struggled to convince their managers, who had concerns about the time it would take and the disclosure of organisational practices. One web developer, who did not take part in the study, revealed that his manager had responded: "If they want our information, they can bloody well pay for it." Because of these difficulties, the representativeness of the participants may have been compromised. Fortunately, I was able to recruit many participants from i2Web consortium organisations. Though I had concerns that the involvement of these participants in the i2Web project might have heightened their awareness of web accessibility, this did not appear to be borne out in the findings of this study.

The web developers in this study could, understandably, only spare a limited amount of time. To increase the likelihood of observing a broad range of web development activities within the limited time available, I prompted web developers to carry out certain tasks rather than wait for those tasks to naturally occur. Beyer and Holtzblatt (1997) argue that prompting participants in this way generates more relevant data, which outweighs any interruption to their natural workflow. Though slightly more artificial, this more directed approach to contextual inquiry ensured minimal disruption to the participant's working day and allowed me to focus on the aspects of their work that were most relevant to the investigation. It also forced the web developers to provide evidence to support their claims. The benefit of this approach became apparent when one web developer in this study, who claimed to regularly use an automated accessibility tool, discovered that the tool had actually been discontinued for several years and no longer worked.

The contextual design methodology is a flexible set of techniques that have been found to be successful in exploring working practices in a wide variety of contexts, from emergency medical despatch (Furniss & Blandford, 2007) to air traffic control (Feigh, 2006). Despite this apparent flexibility, the contextual design methodology has some limitations when applied to modern 'digital workplaces', such as those of web developers, where the work frequently involves mainly internal mental processes and is largely conducted at a computer via a user interface. These constraints often make it very difficult to observe and understand what people are doing without disrupting and potentially influencing their work. Furthermore, some aspects of the contextual design methodology seem particularly difficult to apply to this type of work. For example, it may be impossible to accurately represent the physical environment or workplace culture of a self-employed web developer working out of a café or library. The contextual design methodology–and in particular the five work models–has an inherent focus on the practical, concrete aspects of work, which can result in the more abstract characteristics being overlooked. This was particularly evident in the current study, in which none of the five work models captured the web developers' lack of knowledge and confidence about web accessibility: a big cause of their struggle to create accessible websites. The adoption of a more interruptive approach to contextual inquiry, and the focus upon 'virtual' as well as 'physical' artefacts helped to mitigate the challenges, but further investigations of this nature would benefit from efforts to define and refine methods for collecting, analysing and representing non-traditional, computer-based work. Nevertheless, the contextual design methodology provided a very detailed and revealing snapshot of this particular group of people and highlighted a number of opportunities for improving and supporting their working practices.

4.5. Conclusions

This chapter has presented a contextual inquiry investigation into the working practices of 13 professional web developers. The aim of this study was to examine the role of professional web developers and their working practices in much greater detail to understand why they are struggling to make websites accessible and identify how best to support them.

The findings indicate that web developers want to create accessible websites and learn about the people who benefit from web accessibility. Despite this enthusiasm, web developers lack not only the necessary time and budget to make websites accessible but also the knowledge and confidence to make accurate assessments of accessibility. Though better web accessibility evaluation tools may go some way towards supporting web developers, they appear to contribute little to web developers' understanding of web accessibility and their indiscriminate application may even promote a false sense of confidence in some web developers. Similarly, though web accessibility experts may absolve web developers of the responsibility for making websites accessible, they may also remove the impetus for web developers to further their own knowledge of web accessibility.

To avoid the potential drawbacks of such solutions, web developers instead need access to information that will give them the knowledge and confidence to make such assessments themselves. This information needs to comprise precise, concrete and testable statements that draw upon the language and terminology with which web developers are familiar and provide them with the reasoning behind any technical guidance. Having access to relatable, relevant and rigorous information will further web developers' knowledge of web accessibility, encourage them to make informed accessibility assessments of their websites, and allow them to more easily incorporate web accessibility into their existing web development workflows.

The contextual design methodology has produced a considerable amount of detailed information about the working practices and accessibility requirements of web developers that had not been elicited in previous surveys and interviews. The method has proven extremely useful for understanding the needs of stakeholders in the accessibility chain and establishing a set of requirements for providing accessibility support. Despite the strong focus of the contextual design methodology upon working practices, it has provided a means of understanding the difficulties that web developers face in attempting to make their websites accessible. To more comprehensively understand the nature of web developers' confusion over web accessibility, it is necessary to explore their understanding of web accessibility in much greater detail and, in particular, look for further evidence of a 'developer mindset'. This will provide the basis on which to deliver more relevant and appropriate accessibility support to web developers.

Chapter 5. A study of web developers' mental models of web accessibility

5.1. Introduction

In the contextual inquiry study, I identified some of the problems that web developers encounter in attempting to integrate web accessibility into their workflows. Despite being genuinely interested in web accessibility, web developers still struggle to develop accessible websites. Web developers are hindered, not by limited awareness or concern, but by a lack of knowledge and practical guidance on how to make websites accessible. Existing tools, guidelines and resources let web developers down by not providing them with the support and information they need. Three key themes emerged as to why web developers struggle with existing tools, guidelines and resources: language and terminology; organisation and structure; and volume and comprehensiveness.

While the contextual inquiry study provided an insight into the working practices of web developers, it revealed little about their knowledge and understanding of web accessibility. One participant hinted at the existence of a "developer mindset" that is at odds with current approaches to web accessibility. However, the nature of this confusion is unclear: what do web developers understand about web accessibility? Do they have an appropriate mental model of the subject?

This chapter presents an interview study with 26 professional web developers aimed at eliciting their mental models of web accessibility. It explores what web developers understand about web accessibility and looks to identify any knowledge gaps, inconsistencies and uncertainties that may result in inaccessible websites. It also seeks to uncover any misconceptions or faulty thinking that informs web developers' current approaches to web accessibility.

5.2. Literature review

5.2.1. What are mental models?

Mental models are commonly understood to be internal cognitive frameworks that enable us to understand how something works in the real world. They have been found to be an alluring concept across many fields of research, including: philosophy, education, psychology, cognitive science, and HCI. Researchers have appealed to the concept of mental models in areas as varied as the human mistakes that led to the Chernobyl nuclear disaster (Johnson-Laird, 1999) and people's understanding of home central heating systems (Kempton, 1986). Conant and Ashby (1970) consider the development of these internal representations of external situations to be a "compulsory" component of cognition. The idea that the brain models its environment is, they claim, unquestionable. Similarly, Rouse and Morris (1986) contend that without resorting to the concept of mental models, many aspects of human behaviour are difficult to understand. Payne, Squibb and Howes (1990) consider mental models to be "one of the most important ideas in the psychology of human-computer interaction" (p. 417).

The concept of mental models has endured for a very long time. German physicist, Heinrich Rudolf Hertz (1899) proposed that humans form "internal images or symbols of external objects" that underlie our thought processes and allow us to simulate the external world. Around the same time, American philosopher, Charles Sanders Peirce (as cited in Johnson-Laird, 2004) claimed that humans form internal "diagrams" on which we perform "mental experiments" to reason about the world around us. The picture theory of language proposed by Austrian philosopher, Ludwig Wittgenstein (as cited in Johnson-Laird, 2004), proposed that humans use language to form "pictures of facts" corresponding to reality. Similarly, French philosopher, Georges-Henri Luquet (1927, translated 2001) argued that children construct "internal models" through interaction with the external world. The term "mental model" was not actually coined, however, until 1943 when Scottish philosopher and psychologist, Kenneth Craik used it to refer to the "small-scale models of external reality" that humans construct to explain, to reason, and to anticipate future events (Craik, 1943).

Despite the enduring popularity and apparent necessity of mental models, there is still a surprising lack of consensus among researchers about what constitutes a mental model. This is apparent from many reviews of the mental models literature (e.g. Rouse & Morris, 1986; Wilson & Rutherford, 1989; Moray, 1999; Payne, 2007; Turner & Sobolewska, 2009), which are prefaced with warnings about terminological and methodological inconsistencies. This confusion over the term is notably characterised by Rips (1986) as "mental muddles." Richardson and Ball (2009) attribute the source of the confusion to the plethora of terminology afflicting the literature on mental models.

They highlight how mental models are referred to, often interchangeably, by a range of terms, including: mental representations, internal representations, mental images, conceptualisations, mental simulations and many more. They also comment that "the lack of unified terminology is confusing, especially when the same type of model is described with different vocabulary, by different authors, from different disciplines and from different perspectives" (Richardson and Ball, 2009, p. 337). Payne (1992), on the other hand, considers the confusion over mental models to have arisen from different authors using a single common term to represent very different ideas. Payne (2003) attributes the source of much of this confusion to the 1983 publication of two very different books with the same title: Mental Models.

In one of these books (Johnson-Laird, 1983), mental models represent the cognitive processes underpinning people's thinking, reasoning and problem solving. Mental models describe the mental mechanisms that allow people to process information, make inferences and deduce conclusions. For example, given the premises "Person A is taller than Person B" and "Person C is shorter than Person B", Johnson-Laird alleges that we form an analogous mental representation of the three people allowing us to deduce the conclusion that Person A is taller than Person C. Similarly, given the premises "The scissors are to the left of the stone" and "The paper is to the left of the scissors", we allegedly form and consult a mental model of the spatial arrangement. The construct of such models allows us to make deductive inferences (e.g. that the paper is to the left of the stone) that were not explicitly mentioned in the original premises. These representations are believed to be transient and dynamic and manipulated 'on the fly' within people's working memory (Wilson & Rutherford, 1989). This interpretation of mental models, as the dynamic cognitive structures that allow people to process information and reason about the world around them, has also been extremely influential and inspired a substantial body of research.

An alternative interpretation of mental models is presented in the other book published in the same year (Gentner & Stevens, 1983), in which mental models represent the collection of common-sense knowledge, beliefs and theories that people formulate about the world. Drawing on a disparate array of domain-specific examples, including motion, mechanics, ocean navigation, electricity, calculators and computer systems, Gentner and Stevens characterise such mental models as "naive representations" that humans construct to guide their thinking and behaviour. For example, many people adopt the relatively simple analogy of water flowing through pipes to understand the more complex concept of electrical current "flowing" through wires. Similarly, many people liken the behaviour of an atom to that of the solar system with electrons orbiting the nucleus in the same way planets orbit the sun. Such analogies may not always be accurate (and in many instances may be completely wrong) but they provide people with a working knowledge on which to base their actions. These static and relatively unchanging representations (often referred to as conceptual models) are believed to be constructed within people's long term memory (Welford, 1961). This interpretation of mental models, as domain-specific knowledge that allows people to reason and make inferences about that particular domain, has also been very influential, particularly in the field of HCI, where mental models have been used to represent people's conceptualisations of computer systems and devices.

Despite contrasting interpretations of mental models and the plethora of terminology used to label and describe them, the field of mental models may not be as conflicted and uncertain as it appears. Johnson-Laird (1989) suggests that different approaches to mental models merely represent different aspects of the same cognitive phenomenon, claiming that "theorists are talking about the same beast" (p. 490). This is echoed in Richardson and Ball's (2009) comprehensive review of the mental models literature, which begins by reaffirming the seemingly contrasting interpretations of mental models proposed by Johnson-Laird (1983) and Gentner and Stevens (1983). The review concludes, however, that the two interpretations are, in fact, complementary. Together, they form a cognitive framework in which dynamic representations constructed in working memory during a particular activity are populated and informed by static representations of knowledge and theories residing in long term memory. In other words, people hold knowledge, theories and beliefs about the world around them and use this to support the mental mechanisms that allow them to process information, make inferences and deduce conclusions. Thus, the two types of cognitive representation can both be considered mental models, working together to underpin all higher-level cognitive activity, including thinking, reasoning, and problem solving.

5.2.2. Mental models and human-computer interaction

Mental models have found enduring application in the field of HCI, where they are used to represent people's understanding of devices and interactive systems. The concept has been surprisingly popular and has received considerable attention, not only from academics in the field but also from industry practitioners. Academics tend to be interested in how people acquire, process and apply knowledge of devices and interactive systems (Bibby, 1992), whereas industry practitioners typically draw upon such knowledge to inform user interface design (Payne, 2007). This prevailing interest has generated a substantial body of academic literature and has led to the concept of mental models being incorporated into industry style guides and user interface guidelines (e.g. the IBM Common User Access Guidelines, 1992; the Microsoft Inductive User Interface Guidelines□, 2001; and the Apple Human Interface Guidelines□, 2017). While the role and purpose of mental models is generally agreed upon in HCI, their precise nature is contested. A similar lack of consensus and variety of terminology to that found in the broader study of mental models afflicts mental models research in HCI. In a contribution to Gentner and Stevens' (1983) influential anthology on mental models, Young (1983) tentatively collates the various theories of what constitutes a mental model of an interactive device, acknowledging that "there are probably as many different ideas about what it might be as there people writing about it" (p. 35).

Despite confusion over the nature and nomenclature of mental models in HCI, they are generally understood to comprise two components: a *conceptual* understanding of a device or interactive system's internal functioning, and a *procedural* awareness of how that internal functioning corresponds to "real world" tasks and activities (Wilson & Rutherford, 1989). This distinction, between our mental representations of the "real world" task domain and the "internal world" device domain, is a well-established paradigm in the field of HCI (e.g. Norman's (1988) human action cycle). It may also be unique to the field: Richardson and Ball (2009) point out how other domains, such as (to use their example) object assembly, do not require an understanding of the inner workings of artefacts.

One of the first authors to make the distinction was Young (1981), who, in a study of pocket calculators, identified two types of mental model: a surrogate model of the internal functioning of a device and a model for mapping tasks onto actions. Young's surrogate models are similar to the "naive representations" proposed by Gentner and Stevens (1983). They provide a conceptual understanding or "cover story" of how devices or interactive systems are believed to work. They may also exploit people's familiarity with existing devices and interactive systems by drawing upon metaphor and analogy. For example, computer operating systems commonly adopt the metaphorical concept of a desktop, complete with files, folders, trash cans or recycling bins, to convey simply their complex internal functioning. Task-action mapping models, on the other hand, are similar to the dynamic representations proposed by Johnson-Laird (1983). They represent how real world tasks and activities (e.g. multiplying the numbers 2 and 3) correspond to procedural actions that can be performed with devices or interactive systems (e.g. pressing the buttons '2', '*', '3' and '=' on a pocket calculator). Payne et al. (1990) consider this procedural/conceptual distinction to have been hugely influential in the study of mental models in HCI.

Task-action mapping models were of particular interest to Payne et al. (1990), who developed a particular type of model, called the Yoked State-Space (YSS) hypothesis. The YSS draws upon the notion of "problem spaces", a concept first introduced in the field of cognitive psychology by Newell and Simon (1972) to analyse problem-solving behaviour. Problem spaces represent a set of states that a problem can be in. Beginning from an initial state, the problem solving process involves mentally searching through the problem space for a sequence of intermediate states that leads to a desired goal state, representing the problem solution. Payne et al. assert that to interact with a device, users must create and maintain two separate problem spaces: a *goal space* and a *device space*. The goal space represents the possible states of the "real world" that the device can manipulate, whereas the device space represents the possible states of the "internal world" of the device. Entities in the device space are semantically mapped or "yoked" to corresponding entities in the goal space. Users may use so-called "device operators" to shift between different states in the device and these shifts correspond to movements in the goal space. Payne et al. (1990) maintain that the complexity of a device, and its ease of use, is reflected in how well the goal space and the device space are yoked: the clearer and more consistent the mappings, the easier users will find the device to use.

5.2.3. Web developers' mental models

Mental models research in the field of HCI focuses primarily on users and their understanding of the devices and interactive systems that they use. Very few studies have examined the mental models of programmers, software engineers and web developers who create and maintain those devices and interactive systems. Programming may be considered a complex problem solving activity. Chevalier, Fouquereau and Vanderdonckt (2009) characterise software and web development as a highly iterative, open-ended, and often ill-defined process that requires developers to rapidly shift between problem structuring and problem solving while constantly evaluating the outcome.

It is reasonable to assume that programmers, software engineers and web developers form some sort of mental model that represents their knowledge of how to actually program a user interface. According to Fischer (1991) and Allen (1997) the design process of any interactive system (be it devices, software applications, or websites) involves the interaction of several different types of mental model. These may include models of the capabilities of the tools; of the partially completed work; and of the users' requirements and capabilities. LaToza, Venolia and DeLine (2006) conclude that software developers go to great lengths to create and maintain rich mental models of code that are rarely explicitly recorded. They found that software developers are able to provide detailed explanations of code architecture, code implementation details, ownership and history of code, without referring to any written material.

In the same way that a pocket calculator user maps real world tasks and activities onto actions that can be performed with the device, programmers, software engineers and web developers must map goals they want to achieve onto coding procedures that allow them to achieve it. Furthermore, their procedural knowledge is likely to be underpinned by a conceptual model that represents the knowledge, theories and beliefs they hold about the user interface, about the tools used to create the users interface, and about the users themselves. Littman, Pinto, Letovsky and Soloway (1987) demonstrate that programmers who have a conceptual understanding of a computer system are more successful in debugging code than those who do not. Inappropriate or incorrect conceptual models may not provide a suitable grounding on which to base their procedural models. Similarly, inappropriate or incorrect procedural models may not feed back to and inform their conceptual models.

The importance of robust task-action mappings that are underpinned by an appropriate conceptual model is further emphasized when it comes to developing *accessible* interactive systems. The extent to which programmers, software engineers and web developers program in an accessible way must depend on whether their conceptual model incorporates the needs and requirements of disabled people.

5.2.4. Eliciting mental models

Perhaps as a consequence of the uncertainty over what exactly constitutes a mental model, the methods that researchers have devised for eliciting and representing mental models are equally eclectic. A footnote in Norman (1983) warns: "discovering what a person's mental model is like is not easily accomplished … you cannot simply go up to the person and ask" (p. 11). Young (1983) offers a broad framework for eliciting and observing users' mental models of interactive systems. He suggests that researchers should observe: users using the system; users explaining the system (including explanations about the causes of an event and diagnoses of the reasons for malfunctions); users predicting the behaviour of the system (predicting what will happen next in a sequential process and how changes in one part of the system will be reflected in other parts of the system); and users learning the system.

Jones, Ross, Lynam, Perez and Leitch (2011) distinguish between direct and indirect mental model elicitation techniques. Direct elicitation techniques require participants themselves to form a representation of their understanding. This is often a diagrammatical representation or spatial arrangement of key concepts, using pictures, words or symbols. Kerr (1990), Hendry and Efthimiadis (2008), and Thatcher and Greyling (1998) used this approach to elicit participants' mental models of, respectively, databases, search engines, and the Internet. LaToza, Venolia and DeLine (2006) used this approach, in addition to surveys and semi-structured interviews, to understand how developers create and maintain mental models of code. Indirect elicitation techniques infer a person's mental model from written documents or verbal descriptions, which may be elicited by qualitative methods, including surveys, interviews, focus groups and contextual inquiry. Ma, Ferguson, Roper and Wood (2007) developed a questionnaire that asked novice programmers to predict the consequences of particular actions (in this example, executing a program). Different responses were mapped onto possible mental model representations.

5.3. Method

5.3.1. Design

The study design combined a number of direct and indirect mental model elicitation techniques from the literature[50].

Firstly, inspired by Young's (1983) advice to observe users explaining the system, participants were asked to describe what they understood by the term "web accessibility". This would provide an opportunity to explore their knowledge of web accessibility, without providing any prompts or cues that might either constrain or inform their response. The question was completely open-ended, and participants were free to provide as brief or extensive a response as they preferred.

Secondly, inspired by Young's (1983) advice to ask users to predict the behaviour of the system, participants were asked to predict a) the web development techniques that they think might benefit different user groups (e.g. blind users or physically disabled users), and b) the users groups that they think might benefit from different web development techniques (e.g. providing subtitles and captioning of multimedia content or ensuring webpages can be resized). As well as exploring the extent to which participants can "run" their mental models to predict the consequences of particular actions, this would also establish the strength of their mappings between real world tasks (i.e. supporting different user groups) and procedural actions (i.e. applying web development techniques). Participants were encouraged to list as many or as few techniques/user groups as they wished without justification.

Thirdly, participants were interviewed about their existing working practices and attitudes towards web accessibility. In addition to further exploring their knowledge of web accessibility within the context of their work, this provided the opportunity to replicate aspects of the contextual inquiry study and to explore the broader applicability of its findings with a larger and different population of web developers.

---

[50] The study was conducted alongside data collection for the i2Web Project. I was assisted in the data collection by two researchers working on the i2Web Project: Anna Bramwell-Dicks and Lucy Buykx. Anna conducted eight of the twenty-six interviews, Lucy conducted five interviews, and I conducted the remaining thirteen interviews. The analysis is entirely my own.

5.3.2. Participants

Twenty-six professional web developers took part in the study. Three participants were female, the rest were male. The age of participants ranged from 22 to 44 years (mean: 31.0). Three participants were from Italy; the remainder were from various parts of the UK. Five participants worked for i2Web consortium organisations. The others were recruited using opportunistic sampling from web developer mailing lists, online forums, social media, conference flyers and personal contacts in the industry. The participants had between 1 and 14 years of experience of web development, with an average of 8 years. Six participants worked for large enterprises (250+ employees), fifteen worked for SMEs (< 250 employees) and five participants were self-employed freelancers.

The study was conducted at the same time as the study presented in Chapter 6. For their participation in both studies, participants were offered £30 worth of Amazon vouchers in compensation for their time and effort.

5.3.3. Materials

A recruitment flyer was distributed at the CHI 2013 conference in Paris, France. At the start of the interview, participants were asked to complete an informed consent form (see Appendix A).

Participants were asked to predict the web development practices and techniques that they think might benefit the following user groups[51]:

a) blind users;
b) partially-sighted users;
c) hard of hearing users (This includes people with a mild or moderate hearing loss. This includes people who have age-related deafness. People who are hard of hearing do not use sign language);

[51] The brief descriptions in parentheses were provided to clarify the distinction between hard of hearing users and deaf users and also to constrain the definition of physically disabled users to people with upper limb disabilities that may affect their use of the web.

d) deaf users (This includes people who have a severe or profound hearing loss. They may need hearing aids and may also rely on lipreading. Sign language may be their first or preferred language);
e) physically disabled users (This specifically refers to people with upper limb disabilities that may affect their use of the web);
f) people with dyslexia; and
g) older adults.

Participants were asked to predict the user groups that they think might benefit from the following web development practices and techniques:

a) provide subtitles and captioning of multimedia content;
b) avoid referring only to the position or colour or shape of an object;
c) provide descriptive headings and titles;
d) ensure web pages can be resized up to 200%;
e) ensure a logical order and grouping of the components on a webpage (form controls, buttons, links, tables etc.);
f) ensure text is of an appropriate size, contrast and justification, with sufficient space between letters, lines and paragraphs;
g) make sure webpages can be tabbed through and operated using a keyboard;
h) provide a way of pausing or stopping rapidly updating content (flashing, scrolling, blinking etc.);
i) include skip links to allow users to jump to the main content on a webpage or to different sections on a webpage; and
j) provide information about a user's location within a website (e.g. using breadcrumb trails, site maps, navigation menus etc.).

The semi-structured interview schedule explored whether the participants currently attempt to make their websites accessible; whether they conduct any accessibility testing and how they go about it; and whether their clients have any accessibility requirements or expectations. The interview schedule can be found in Appendix C.

5.3.4. Procedure

Face-to-face sessions with the participants were arranged at a location convenient to them. In many cases, this was at the participant's workplace, but several interviews were conducted in cafés, restaurants, and at participants' homes. Each session was audio recorded using a mobile device.

Following a brief introduction to the investigation and its aims, participants were familiarised with the interview method. They were assured that the information they would be providing would be confidential and anonymous and their consent was gained to take audio recordings for later transcription. Participants then read and signed an informed consent form. Once consent had been obtained, participants were interviewed about their mental models of web accessibility.

Once the interview was complete and the participant had been debriefed, they were thanked for their time and asked whether they would be interested in taking part in further research. The interviews lasted approximately 30 minutes each.

5.3.5. Data preparation

The audio recordings from each participant were fully transcribed prior to analysis. Sample transcripts from three participants can be found in Appendix G.

Participants' responses to the question "what do you understand by the term 'web accessibility'?" were analysed using a conventional content analysis approach (Hsieh and Shannon, 2005), involving several phases. Firstly, following a close read of each participant's response, I extracted all relevant content words; close synonyms (e.g. "blind people" and "people who are blind") were grouped together. I then recorded the frequency with which each content word was mentioned. Content words that were only mentioned by a single participant were eliminated. From this open coding approach, I derived an initial set of 28 codes. Secondly, I grouped the codes into five broad concept categories (a list of codes and categories can be found in Appendix G). I then combined the concept categories to form an overall hierarchy of codes, representing the participants' aggregate mental model of web accessibility.

Participants' responses to the prediction questions comprised: a) the web development techniques that they felt would benefit each user group, and b) the user groups that they felt would benefit from each web development technique. I scored each participant's responses according to whether they had correctly identified one or more techniques/user groups. To gauge the comprehensiveness of their responses, I recorded the number of techniques/user groups mentioned by each participant. I also noted any comments made by participants during this exercise that I felt reflected their mental model of web accessibility.

The participants' responses to the questions about their working practices and attitudes to web accessibility did not require content analysis. I collated these responses and, where appropriate, calculated descriptive statistics.

Again, conscious of being a sole researcher, and the threat to validity and objectivity that this can pose, I sought to validate my analysis of the initial question—"what do you understand by the term 'web accessibility'?"—with an independent researcher. I did not seek to validate my analyses of the remaining questions, as they were largely descriptive.

Due to the open-ended nature of the initial question, which allowed participants to provide as brief or extensive a response as they preferred, validation proved difficult. The participants' responses were conversational and uneven and, because they did not speak in clearly delineated text units, the portions of text to be analysed (e.g. a phrase, a sentence, a paragraph) were inconsistent. Krippendorff (1995) refers to this as the *unitisation problem*, due to the possibility of different coders unitising the same text differently and disagreeing on which portions of text contain a particular meaning. This makes it hard to determine whether their coding is the same, which, ultimately, makes it difficult to calculate inter-coder reliability (Campbell, Quincy, Osserman and Pedersen, 2013).

Campbell et al. (2013) proposed a solution to this problem by having the primary, more experienced, coder first define and code the units of analysis before having the secondary coder code these specific units. Although this approach achieves greater consistency of unitisation and makes it easier to calculate inter-coder reliability, it introduces the possibility of bias. For example, the predefined units of analysis may alert the secondary coder to the fact that there is something to be coded in a portion of text that they may not have otherwise coded. This approach also fails to address the problem of multiple codes being applicable to a single portion of text, which further complicates the calculation of inter-coder reliability.

In lieu of clear guidance in the literature for coding semi-structured interview transcripts (for a candid discussion of the problem, see Campbell et al., 2013), I did not attempt to calculate inter-coder reliability. Instead, I invited the independent researcher to code three of the participants' responses to the initial exercise (a subset of approximately 10% of the collected data) using the same codes I developed during my analysis. The unit of analysis in this case was each participant's entire response, to which multiple codes could be applied. A degree of consistency between the researcher's and my application of the codes would then serve to corroborate, if not confirm, the results.

As in the contextual inquiry study (see Section 4.2.4), I resolved any disagreements between the two coding attempts (of which there were very few) through a negotiated agreement process. Although not as rigorous as statistical intercoder reliability testing, this approach brought a degree of verification and validation to my analysis of the unstructured interview data.

5.4. Results

5.4.1. Mental models of web accessibility

Participants' responses to the question "what do you understand by the term 'web accessibility'?" varied substantially in terms of completeness and complexity. The concepts that the participants mentioned, however, were fairly consistent, suggesting a largely uniform mental model of web accessibility. As described in the previous section, five broad concept categories emerged from analysis of the participants' responses.

- **Definitions of web accessibility**: statements about the meaning of the term, for example "web accessibility is about making websites accessible to people with disabilities".
- **Beneficiaries of web accessibility**: descriptions of users groups who benefit from web accessibility, for example, blind or partially-sighted people.
- **Technologies relating to web accessibility**: descriptions of both assistive and mainstream technologies, including hardware and software, for example, screen readers or alternative keyboards and mice.
- **Coding practices for implementing web accessibility**: descriptions of techniques and best practices for developing accessible websites, for example, providing alternative text for images or ensuring videos have subtitles.
- **Methods for evaluating web accessibility**: descriptions of accessibility evaluation methods, for example, using an automatic evaluation tools or conducting user testing.

The following is a breakdown of each category.

5.4.1.1. Definitions of web accessibility

Common to each of the 26 participants' responses was a basic definition of web accessibility. Though the complexity and completeness of their responses varied, each participant's definition was reasonably accurate. Furthermore, though their precise terminology varied, the participants' definitions of web accessibility were fairly consistent. Figure 5.1 provides a breakdown of the participants' responses.

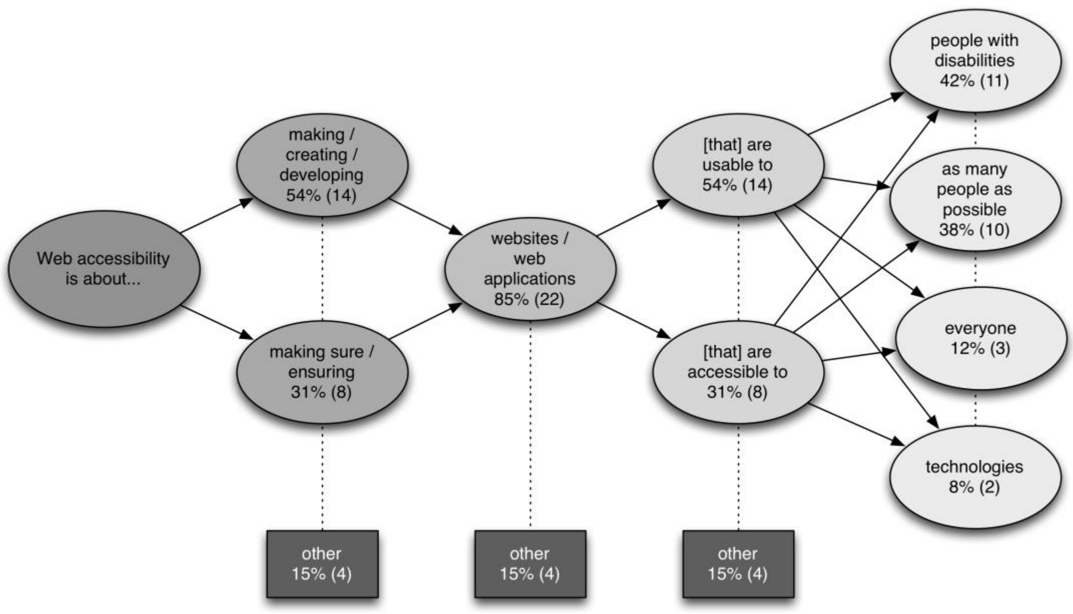


Figure 5.1: Breakdown of participants' definitions of web accessibility

Just over half of the 26 participants (54%, 14) felt that web accessibility is a creative process, drawing upon common terms such as "making", "creating" and "developing" in their definitions. In contrast to this, nearly a third of participants (31%, 8) characterised it more as a process of conformance, using terms such as "making sure" and "ensuring". 15% (4) of participants provided more unique responses, charactering web accessibility as a process of measuring or helping.

Unsurprisingly, almost all of the participants (85%, 22) considered web accessibility to be about websites, web applications or closely related concepts. Other terms used by the remaining 15% (4) of participants included: "interface", "digital products", and "anything that is delivered over the web". One participant referred instead to "any vital, important content" [WD4] rather than a website or web application. Only one participant defined web accessibility without referring to any of the above terminology, stating that it is about: "the number of people who can access something" [WD22].

Just over half of the participants (54%, 14) stated that web accessibility was about making websites or web applications "usable", whereas just under a third of participants (31%, 8) instead used the more obvious term "accessible". This subtle distinction may simply have arisen from participants being reluctant to provide a circular definition of web accessibility, with one noting: "it's going to be very difficult to not say 'accessible'!" [WD22]. Other terms used by the remaining 15% (4) related to websites being "vaguely friendly", "as readable as possible", "suited to the needs" and "seen or watched or understood".

The participants' definitions also varied by the scope of the perceived beneficiaries of web accessibility. The most common response, provided by 42% (11) of participants, was that web accessibility is specifically for people with disabilities. 38% (10) of participants broadened the scope of web accessibility, considering it to be about allowing as broad and diverse a range of people as possible to access or use a website or web application. 12% (3) of participants broadened the scope even further, describing web accessibility as the process of making websites accessible to everyone, irrespectively of their ability, location or technology. This inclusive approach was summarised by one developer who argued that web accessibility should include,

> not only those who have significant challenges through life, more than just the web, but also those with ability and disability, essentially. Perhaps the perfect goal of 100% of anyone who needs to use a website being able to use that website [WD25].

Curiously, two web developers (8%) defined web accessibility in terms of making websites accessible to the ATs used by people with disabilities, rather than the people themselves.

Participants' mental models of web accessibility incorporate a reasonably accurate and consistent definition of the term, characterising it as a user-centred process of developing websites that are usable and accessible to all people, including those with disabilities. However, their choice of verb in describing the process (e.g. "making" or "creating" versus "making sure" or "ensuring") reveals an important conceptual distinction: for some web accessibility is an act of creativity, for others it is an act of conformance.

5.4.1.2. Beneficiaries of web accessibility

As well as providing a definition of web accessibility, the majority of participants (85%, 22) also described specific user groups. Overall, the participants described a range of specific user groups who benefit from accessible websites. However, by far the most frequently mentioned user groups were people with visual disabilities and people with physical disabilities. Figure 5.2 provides a breakdown of the participants' responses.

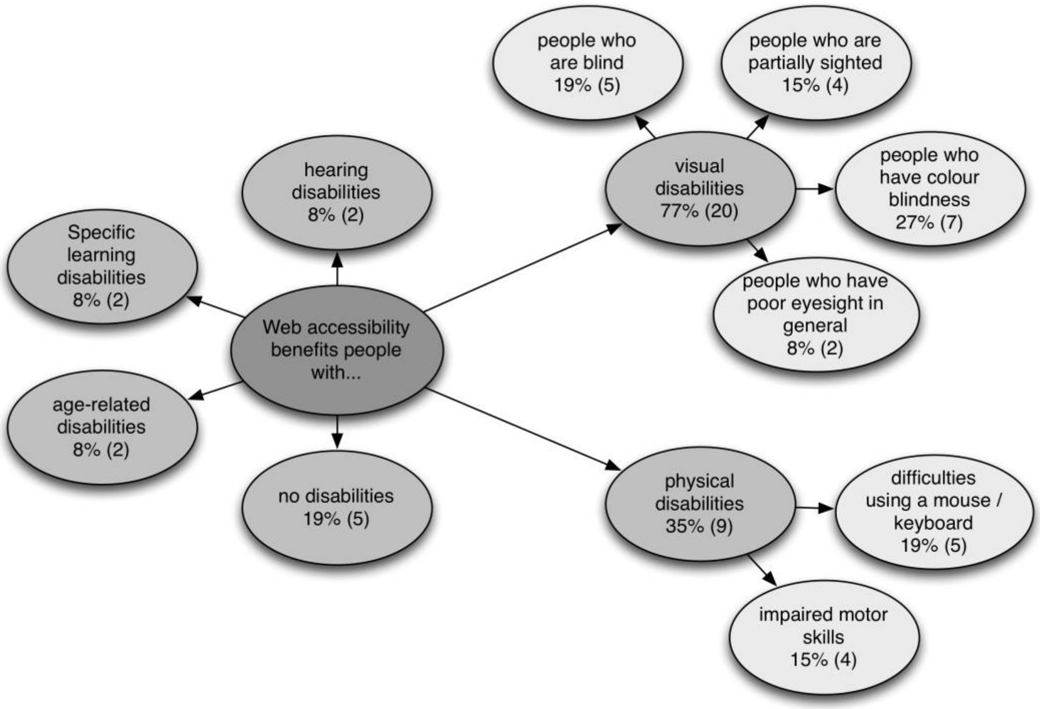


Figure 5.2: Beneficiaries of web accessibility

The majority (77%, 20) of the 26 participants included people with visual disabilities in their responses. Eight of these (31%) included *only* people with visual disabilities in their responses. 19% (5) of participants specifically mentioned blindness or not being able see at all. 15% (4) of participants specifically mentioned partially-sightedness or not being able to see very well. 27% (7) of participants included colour blindness in their responses, and 8% (2) referred to people who have poor eyesight in general.

A number of quotes from participants illustrate the predominance of people with visual disabilities in their mental models. For example, one participant said: "I suppose the main group that springs to mind is sight, so partially-sighted people, colour-blindness are the general groups really" [WD2]. Another participant said: "The people that come to mind are blindness, maybe other visual impairments, so colour blindness, or degrees of blindness and visual impairment, e.g. bad eyesight for some reason" [WD1]. Another participant said: "I had a few people using stuff that I've made where they were blind. I'd say also people who are colour blind. Other than that, I'm not really sure" [WD8]. One participant stated that web accessibility is specifically for: "people that might have an impairment such as someone who is blind or someone who has… [I'm] trying to think of different impairments!" [WD11].

People with physical disabilities were the second most frequently mentioned user group and was included by just over a third (35%, 9) of participants in their responses. 19% (5) of participants specifically referred to difficulties in using a keyboard or mouse. 15% (4) of participants specifically referred to impaired motor skills. None of the participants included *only* people with physical disabilities in their responses. However, 27% (7) of participants included *only* people with visual disabilities and people with physical disabilities, with one participant describing the user groups as "the two big ones" $_{[WD7]}$.

The common co-occurrence of people with visual disabilities and people with physical disabilities in the participants' mental models is illustrated by a number of quotes. For example, one participant said: "I think the main group that we always start with is the visually impaired. And then think about motor skills as well" $_{[WD4]}$. Similarly, another participant said: "when people talk about accessibility, blind people is the one that comes up the most often. People who have physical impairments when using mice and keyboards" $_{[WD18]}$. Another participant stated that web accessibility was: "mostly for sight impaired people and sometimes if you can't use your keyboard and mouse" $_{[WD16]}$. Another participant stated that web accessibility was about making websites "as accessible as possible for a variety of users bearing in mind that not everybody uses a mouse, not everybody uses a keyboard, not everybody is perfectly sighted and so on" $_{[WD22]}$.

Only two participants (8%) included people with hearing disabilities in their responses, one of whom stated that having a hearing impairment is: "a lesser issue" because "you don't necessarily need to hear most things on the web" $_{[WD9]}$. The other participant said: "Obviously we barely use any Flash or anything so we don't need to worry about hearing impaired" $_{[WD16]}$.

Only two participants' (8%) mental models of web accessibility include people with specific learning disabilities. Of these, one participant specifically mentioned dyslexia and the other mentioned dyslexia, dyspraxia *and* dyscalculia. Neither participant referred *only* to people with specific learning disabilities in their responses.

Only two participants (8%) included people with age-related disabilities in their responses, one referring specifically to visual problems encountered by older adults and the other citing older adults' unfamiliarity with technology as opposed to an actual disability.

In accordance with the broader definitions of web accessibility provided by some participants, 19% (5) of participants stressed how everyone can benefit from web accessibility and included people with no disabilities in their responses. Of these, one participant included "end users who won't be very technical" and "people who don't use the language you're presenting [the website] in" $_{[WD15]}$. One participant said: "I also consider [web accessibility] to mean things that are accessible normally, so making sure that type is big enough to read for a normal user, not necessarily someone who has a medical issue" $_{[WD1]}$.

Participants' mental models of web accessibility incorporate a limited awareness of the user groups who benefit from accessible websites. Their focus centres largely on people with visual disabilities and people with physical disabilities: two user groups who encounter very distinct and apparent challenges in accessing the web. The web is often considered to be a visual medium due to how websites are typically rendered in graphical web browsers. This presents a number of challenges of people with visual disabilities. Similarly, the increasingly interactive nature of the web can be particularly challenging for people with physical disabilities. User groups whose access issues are less obvious (such as people with hearing disabilities) are either casually dismissed by the participants or not mentioned at all. The participants' awareness of disabled people may be based on the extent to which people's disabilities appear to limit access to the web. This hints at a narrow conceptual understanding of web accessibility concerned more with facilitating technical access to websites than supporting an inclusive user experience.

5.4.1.3. Technologies relating to web accessibility

As well as providing a definition of web accessibility and describing specific user groups, just over half of the participants (58%) referred to specific technologies (assistive or otherwise) in their description of web accessibility. Some participants did not mention specific technologies but described events that would imply the use of a visual display (e.g. in relation to the appearance of fonts or colour schemes). Figure 5.3 provides a breakdown of the participants' responses.

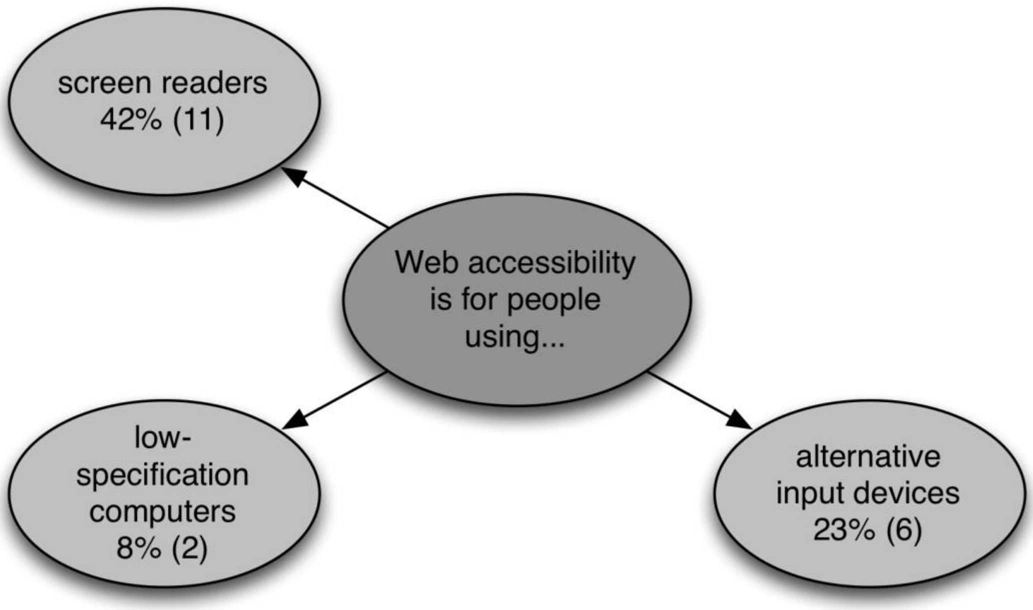


Figure 5.3: Technologies relating to web accessibility

The most frequently mentioned technology was screen readers, mentioned by 42% (11) of participants. This is in accordance with people with visual disabilities being the most frequently mentioned user group. Interestingly, none of the web developers referred to any specific screen reader software (e.g. JAWS or VoiceOver).

The prominence of screen readers in the participants' mental models is illustrated in a number of quotes. For example, one participant said: "accessibility involves making sure your website can be browsed by different types of user agents, e.g. a screen reader for people with a visual impairment" $_{[WD1]}$. Another said web accessibility: "is about getting technology right, so it's things like making sure content is available to screen readers or content locked away in proprietary technologies like Flash is available as plain text. Doing it wrong is not doing that stuff" $_{[WD14]}$.

Six participants (23%) acknowledged the need for websites to work across a range of input devices and user agents and to be "technology neutral". This was not limited only to ATs, with two participants referring specifically to keyboards, and another to mobile phones.

The participants' emphasis on providing compatibility across different technologies is illustrated by a number of quotes. For example, one participant said: "If I was using a mouse, I might expect someone with impaired motor skills to not have access to a mouse and be using another input device" [WD10]. One participant stated that web accessibility is: "generally just making the website vaguely friendly to screen readers and technologies to actually amplify the web experience for impaired people" [WD16]. Another said: "You're making sure that things are keyboard accessible and accessible to screen readers and accessible to people who are visually impaired or colour-blindness or whatever" [WD13]. Another participant felt that web accessibility is "the idea of semantically building sites such that they are more accessible for different tools, different systems and ways of using the web" [WD8].

Two participants (8%) felt that it was important to consider people who might be using older technology, with one participant citing the example of "someone who is in a poorer country who might not necessarily be using a massive great screen with the very latest browser" [WD24]. The other participant argued that the only technological requirement to access the web should be a computer and an Internet connection, adding: "if you have access to those, [web accessibility is] this egalitarian idea that anyone with access to those basic technologies should, as far as possible, be equally able to use the services that are available through them" [WD25].

Participants' mental models of web accessibility incorporate a limited awareness of the ATs used by people with disabilities. Focusing mainly on screen readers and alternative input devices, their knowledge of ATs appears to reflect their limited awareness of disabled people. What is particularly revealing, however, is that many participants describe web accessibility in terms of supporting different user agents but not in terms of supporting actual users. This, again, suggests a conceptual understanding of web accessibility that prioritises technical requirements over user needs.

5.4.1.4. Coding practices for implementing web accessibility

Two participants (8%) felt that web accessibility does not require any particular accessible coding practices. One argued that web accessibility is "generally covered by the way you build it anyway" [WD4]. The other stated that that to create an accessible website you "don't need anything, just programming" [WD21]. However, a large proportion (69%, 18) of participants described specific coding practices for creating accessible websites. Figure 5.4 provides a breakdown of the participants' responses.

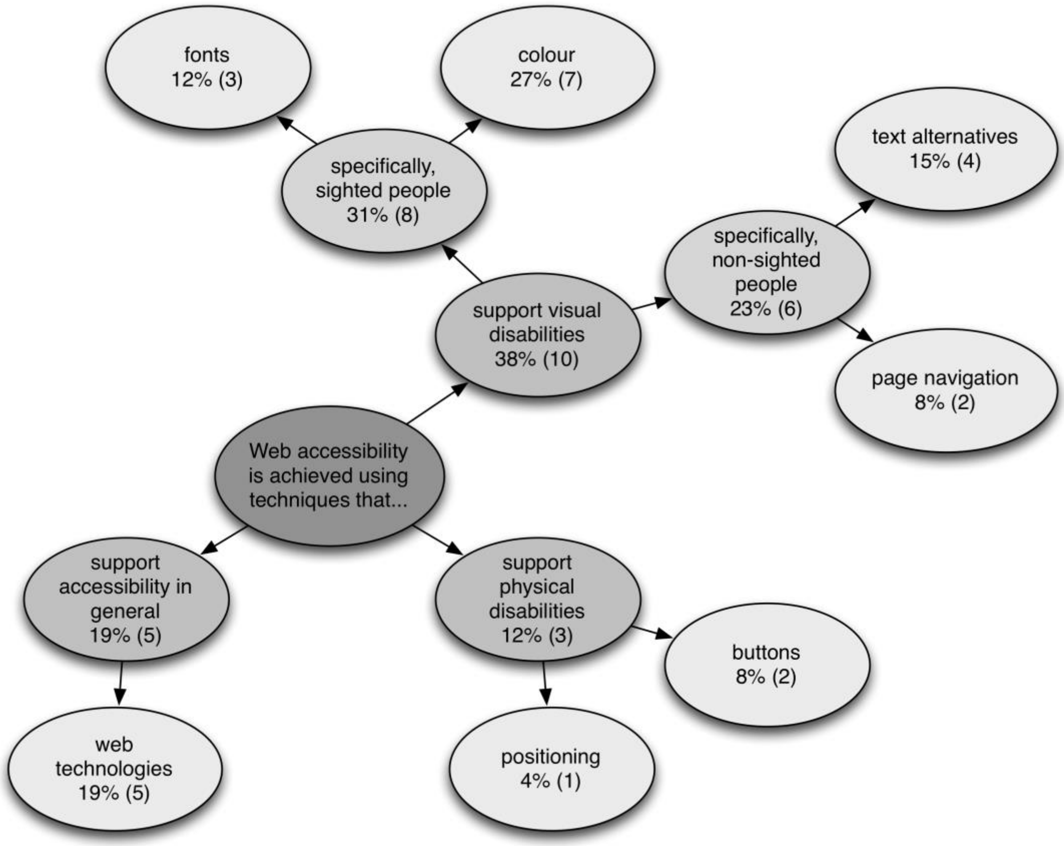


Figure 5.4: Coding practices for implementing web accessibility

Over a third of the participants (38%, 10) mentioned coding practices to support people with visual disabilities. These were further categorised into coding practices that support people that have some sight (mentioned by 31% (8) of participants) and coding practices that support people with no sight at all (mentioned by 23% (6) of participants).

Coding practices that support people with some sight were categorised further still into coding practices related to colours (mentioned by 27% (7) of participants) and coding practices relating to fonts (mentioned by 12% (3) of participants). The coding practices relating to colours included: providing appropriate colour schemes; ensuring webpages have sufficient colour contrast; and avoiding relying on colour to convey information. The coding practices relating to fonts included: making fonts large enough to read; and providing a means of allowing users to adjust the font size.

One participant said: "For me [web accessibility] is mostly stuff like... colours on screen is a big concern, so thinking about how things display to people with different types of colour blindness" [WD7]. In contrast, one participant felt that colour scheme decisions were the responsibility of a web developer, arguing: "I suppose the designer would design it using colours that aren't unsuitable for colour blind people. Obviously, that's more the designer's responsibility so not greens and blues or whatever" [WD2].

Coding practices that support people with no sight at all were categorised further still into coding practices relating to text alternatives (mentioned by 15% (4) of participants) and coding practices relating to page navigation (mentioned by 8% (2) of participants). The coding practices relating to text alternatives included: providing text alternatives for images; and providing text alternatives for Flash components. The coding practices relating to page navigation included: providing a meaningful order of information; and providing skip links at the top of the webpage.

Participants frequently described their efforts in terms of supporting ATs as opposed to enabling users. One participant said: "I try and follow what I believe to currently be best practices, in terms of ordering things so that screen readers can pull out the most important information" [WD20]. Another said:

> I think screen readers look for different tags and that kind of thing and structure and conforming to that helps make things better. One of the simplest things is if semantically your navigation is at the top of the code, then having an invisible link that says 'skip to content' so that screen readers see that first, and you can use that to skip to content so you don't have the whole navigation read out [WD8].

12% (3) of participants suggested coding practices to support people with physical disabilities. These were further categorised into coding practices relating to buttons (mentioned by 8% (2) of participants) and coding practices relating to positioning (mentioned by 4% (1) of participants). The coding practices relating to buttons included making buttons sufficiently large enough. The coding practices relating to positioning included positioning elements to minimise mouse movements.

One participant stated that web accessibility involved: "making sure buttons aren't too fiddly and thinking about how you could operate a site without using a mouse" [WD4]. Another said: "If you've got to do some arduous task to move the mouse across the screen you might want things like being able to use the tab key, rather than having to click in every box" [WD9]. One participant argued that button design is not the responsibility of the web developer, stating: "it is more the responsibility of a designer to make sure buttons are big enough to be visible" [WD2].

Another set of coding practices suggested by 19% (5) of participants did not relate to a particular user group but instead to support web accessibility in general. Each of these coding practices related to the use of web technologies, including: using semantic HTML; using CSS; using web development best practices; and ensuring webpages conform to web standards.

The prominence of standards conformance in the participants' mental models is illustrated in a number of quotes. One participant said web accessibility involves,

> making sure that [a webpage] is actually sensibly marked up. It actually meets standards, that you don't abuse tags, like put everything in paragraph tags for example. Actually mark it up so it can be read without the style sheet or anything attached to it [WD16].

According to another participant, web accessibility "is a wide-ranged approach depending on how you write your markup, whether that's semantic" [WD26]. One participant stated:

> As a developer I would have two definitions [of web accessibility]: one is more related to me, which is web standards: does the website validate correctly, are you using correct HTML and all that stuff. And the other one is can people use it effectively [WD2].

Participants' mental models of web accessibility incorporate a limited procedural understanding of how to implement accessible websites, which reflects their limited knowledge both of disabled people and the ATs they use. Despite referring to a range of accessible coding practices in their responses, participants appear to be most familiar with those that support people with visual disabilities and people with physical disabilities. Notably, participants describe coding websites in a particular way to support screen readers and ATs rather than the people who use them. This again indicates a conceptual understanding of web accessibility that focuses more on technical access than user experience. Also, participants' emphasis on standards conformance, while potentially beneficial to all web users, again indicates a conformance-oriented conceptual understanding of web accessibility.

5.4.1.5. Methods for evaluating web accessibility

Only six participants (23%) made any reference to evaluation methods in their description of web accessibility. Figure 5.5 provides a breakdown of the participants' responses.

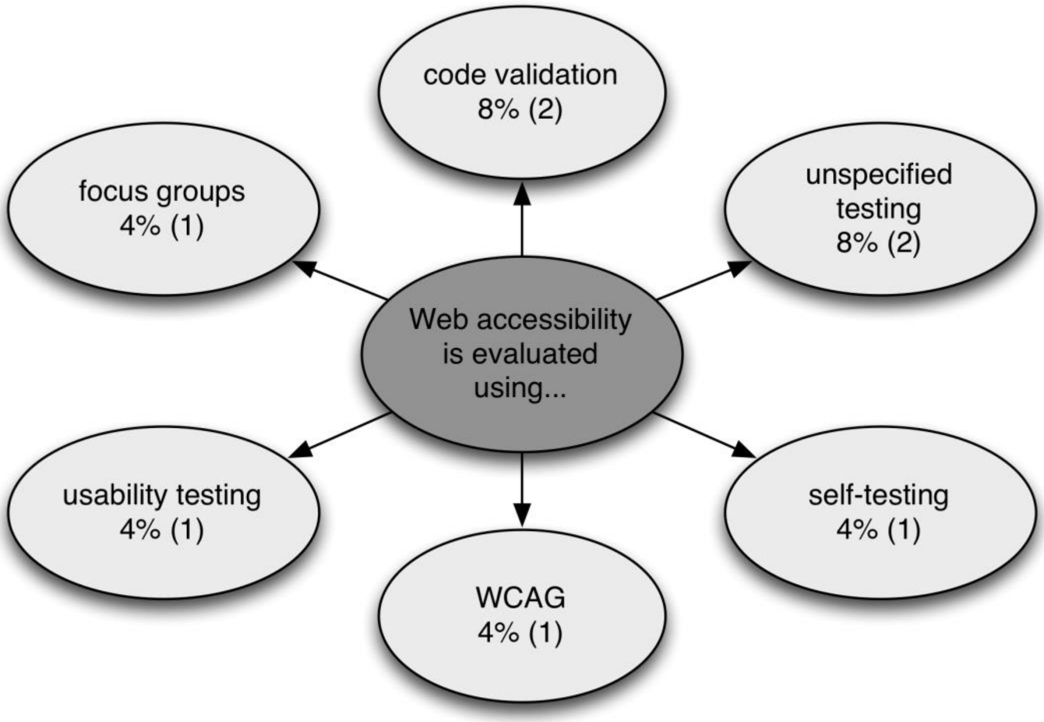


Figure 5.5: Methods for evaluating web accessibility

Two participants (8%) agreed that web accessibility evaluation is necessary but were unable to describe any evaluation methods. One stated that they were unsure of exactly what needs to be tested, as they had "never really done it" [WD15]. The other admitted:

> I've never actually used a screen reader or tried it out to see what it's like, so [web accessibility] is almost stuff you only know from seeing it written about or that kind of thing rather than actually stuff you've seen yourself to see what it's actually like [WD8].

One participant (4%) claimed that usability testing could be used to test the accessibility of websites and another participant (4%) felt that focus groups were an appropriate evaluation method. Notably, none of the participants described testing websites specifically with disabled people. One participant (4%) felt that testing websites by viewing them in a web browser is a sufficient means of evaluating web accessibility, whereas another two participants (8%) stated that running websites through the W3C

HTML Validation Service is sufficient. Only one participant (4%) mentioned the use of WCAG to evaluate the accessibility of websites and web applications but insisted on breaking off the interview to check the name of it.

Of the small number of participants who recognise the need to evaluate the accessibility of websites, their mental models of web accessibility incorporate a sparse and haphazard awareness of how to go about it.

5.4.2. Predictions from mental models of web accessibility

5.4.2.1. Predictions of web development techniques

Figure 5.6 describes the proportion of participants who accurately predicted one or more techniques that would benefit each user group, as well as descriptive statistics of the techniques they predicted.

The majority—between 73% (19) and 100% (26)—of participants accurately predicted at least one technique for each user group, including for those whom they had not previously mentioned (such as deaf and hard of hearing users). This suggests their task-action mappings are relatively strong, allowing them to "run" their mental models to predict the consequences of particular actions, even for less familiar stimuli.

However, the participants' limited awareness of certain user groups is reflected in the number of techniques they mentioned. For blind users and physically disabled users, with whom participants appeared more familiar in the previous section, the mean number of techniques mentioned was 5.2 and 3.5, respectively. For each of the other user groups, however, the mean number of techniques was slightly smaller (1.3–2.6).

Comments made by participants during this question further indicate a limited awareness of certain user groups. For example, eight participants (31%) consider techniques to support older adults to include everything that supports all of the other user groups, whereas one participant (4%) feels there is no need to do anything to support older adults. Three participants (12%) believe there are few or no problems to address for either deaf or hard of hearing users. Six participants (23%) attempted to justify their lack of awareness of techniques to support deaf and hard of hearing users by stating that none of the websites they develop incorporate audio.

The language that participants' use in their responses again reveals a technology-oriented conceptual understanding of web accessibility. They describe ensuring websites "can be browsed by screen readers" or "are keyboard accessible". They also recommend being "aware of alternative input devices" or avoiding "mouse-specific functionality". However, none of their responses refer specifically to users or the actions they perform. Furthermore, many of the techniques mentioned, such as "use appropriate, semantic HTML" or "use 'em' instead of 'px' when defining font sizes", are synonymous with WCAG guidance, which also indicates a conformance-oriented conceptual understanding of web accessibility.

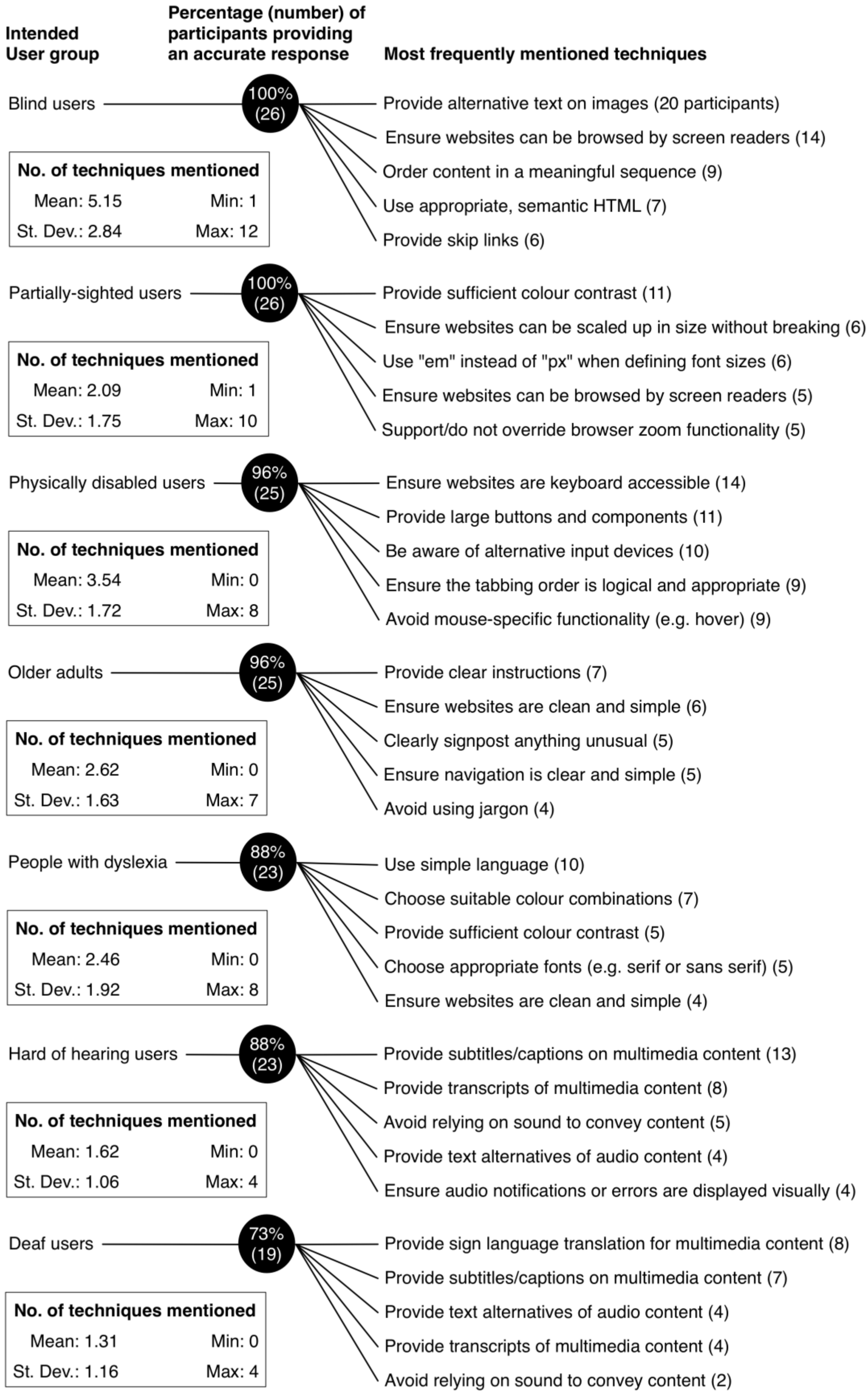


Figure 5.6: Predictions of web development techniques

5.4.2.2. Predictions of user groups

Figure 5.7 and Figure 5.8 describe the proportion of participants who accurately predicted one or more user groups who would benefit from each web development technique, as well as descriptive statistics of the user groups they predicted.

The majority—between 85% (22) and 100% (26) —of participants accurately predicted at least one user group for each web development technique. This included techniques that they had not previously mentioned (such as providing a way of pausing or stopping rapidly updating content). Again, this suggests participants' mental models of web accessibility are "runnable", allowing them to derive accurate outputs, even from less familiar inputs. It also suggests that the "real world" tasks of supporting different user groups are adequately mapped or yoked to the application of different web development techniques.

The dominance of a particular user group or technique is less apparent in this set of predictions. The mean number of user groups mentioned was similar for each technique. However, notice how participants regard blind people or people with visual impairments to be potential beneficiaries of every technique. Although not incorrect, it reaffirms their familiarity with certain user groups. Also, consistent with the more inclusive definitions of web accessibility that participants provided in the previous section, some participants acknowledged that a number of the suggested techniques ultimately benefit everyone.

As in the previous set of predictions, participants' technology-oriented conceptual understanding of web accessibility is revealed by their choice of language, often defining user groups not in terms of their disability but in terms of the ATs they use. For example, the most frequently mentioned user group that participants felt would benefit from the provision of skip links was "screen reader users". Similarly, "keyboard users" and "non-mouse users" were frequently mentioned beneficiaries of ensuring that webpages can be tabbed through using a keyboard. While such responses are technically accurate and may even represent shorthand for the variety of user groups who interact with websites using only a keyboard, they again suggest that participants think more in terms of supporting technologies than supporting users.

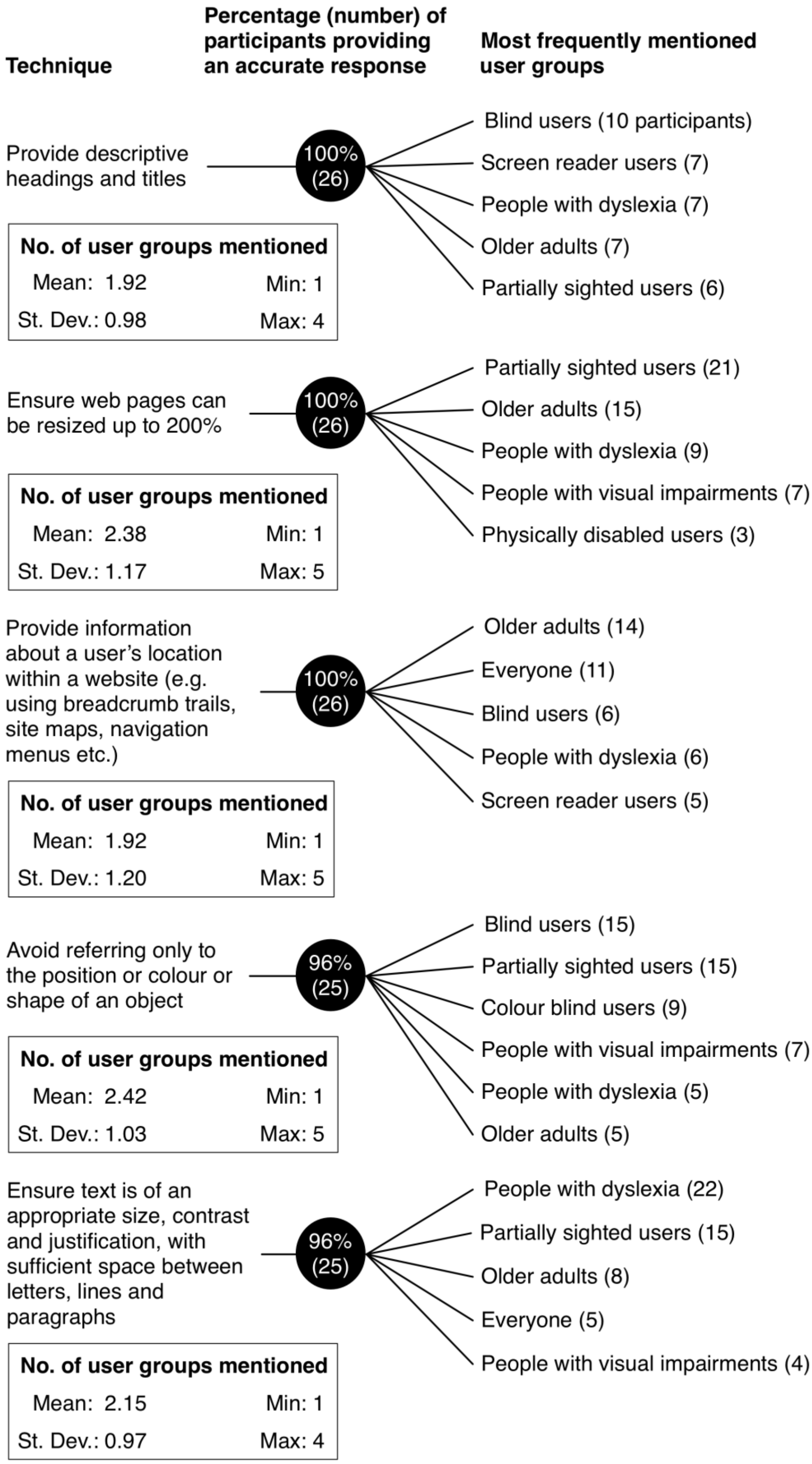


Figure 5.7: Predictions of user groups (1 of 2)

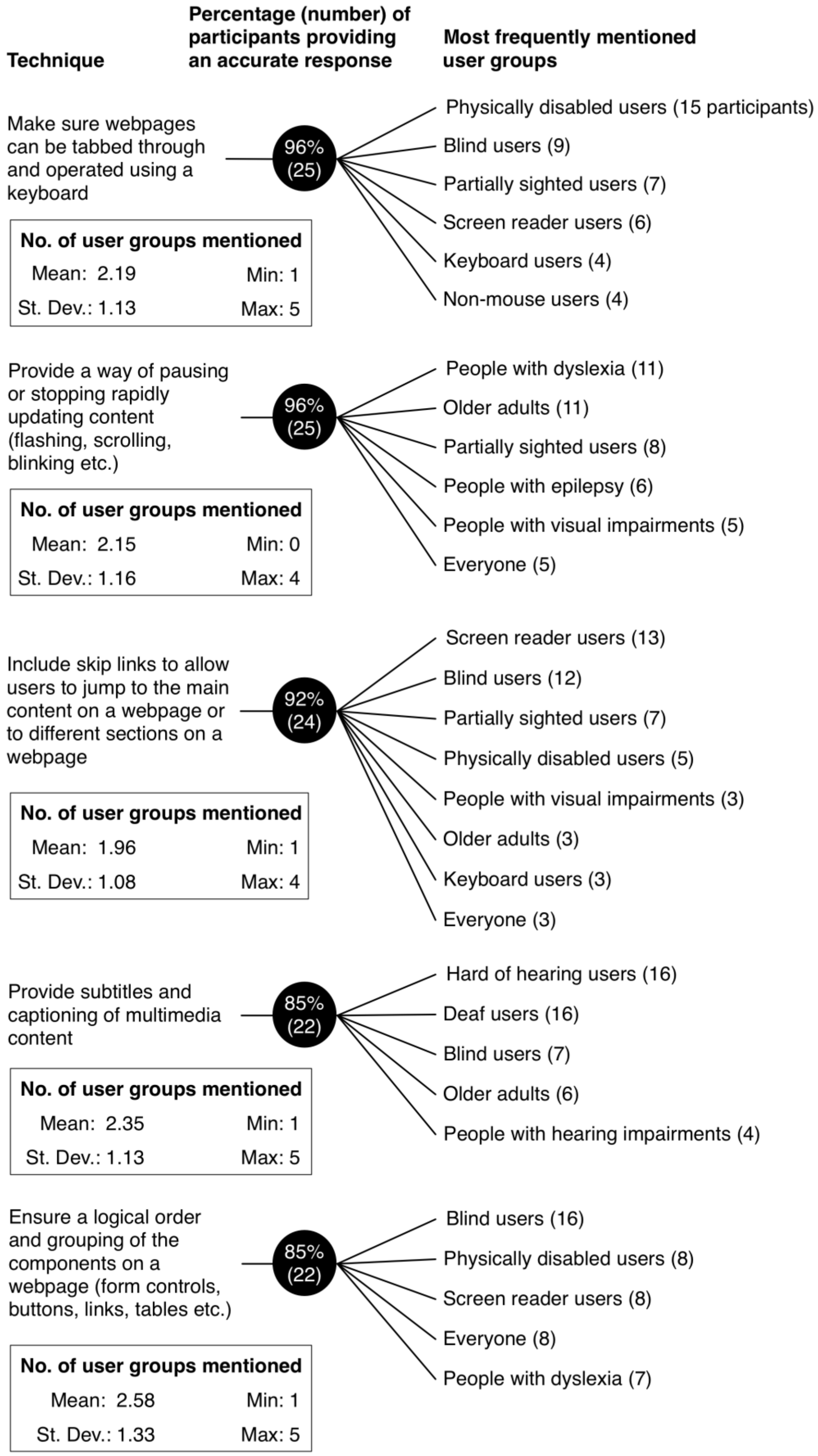


Figure 5.8: Predictions of user groups (2 of 2)

5.4.3. Web accessibility working practices

5.4.3.1. Web accessibility implementation

Participants were asked whether they currently attempt to make their websites accessible. Although four of the 26 participants (15%) admit making no attempt whatsoever to make their websites accessible, the remaining 22 participants (85%) claim to make at least some attempt. Many participants acknowledge not attempting to make websites accessible as often as they should.

Participants who currently attempt to make their websites accessible were asked to describe the sorts of things they do to achieve this. As with their responses to the previous mental models questions, participants appear to be familiar with a range of accessible coding practices. However, many participants admit that, due to constraints upon their time, they rarely go beyond the basics.

The most frequently reported practices involved keeping websites clear and simple by using well-formed, semantic code and reducing the amount of unnecessary content. Many participants admit tackling only the "easy stuff", the "general stuff" or "basic things", such as providing alternative text for images; choosing appropriate colour schemes; and marking up content using semantic elements such as headings, paragraphs and lists. Although not very ambitious, this approach at least hints at an awareness of more advanced techniques.

Some participants admit that the techniques they employ are motivated more by a desire to write clean, efficient code than to produce accessible websites. Others make the dubious claim that web accessibility is an inevitable by-product of other web development practices, such as standards conformance or SEO.

Consistent with the contextual inquiry study, web developers' limited efforts in creating accessible websites—and their limited mental models of web accessibility—appear to stem not from a lack of awareness or concern but from organisational constraints upon their time. Accessible coding practices are rarely implemented proactively but instead tend to emerge as a consequence of other web development practices.

5.4.3.2. Motivations and barriers to web accessibility

Participants were asked what factors either motivate them to or prevent them from taking web accessibility into account. Some claim to be motivated by practical benefits, such as easier maintenance in future. Some claim to be motivated by client requirements or the expectations of their employer. One participant (4%) claims to be motivated by the potential for legal action. The majority of participants however, claim to be motivated by professional pride or a desire to "do the right thing". This is consistent with the contextual inquiry study, which found that ensuring websites are accessible by as many people as possible provides considerable motivation to many web developers.

Predictably, client needs and requirements were the most commonly cited obstacle preventing web developers from making websites accessible. Many participants claim to have neither the time nor the budget to make websites accessible, especially considering the tight development schedules imposed by their organisations. One participant (4%) claims the dynamic, cutting-edge nature of the websites he develops made it difficult to make them accessible. Conversely, another participant (4%) describes working with old websites and "legacy code" which, he claims, would require too much work to make accessible.

According to one participant (4%), being involved mainly in back-end web development absolves him of having to make websites accessible. Another participant (4%) claims the difficulty in verifying the accessibility of the websites he develops makes him feel "powerless" to help people with disabilities. As in the contextual inquiry study, some web developers evidently lack confidence in their ability to accurately assess the accessibility of websites.

While a variety of factors motivate web developers to take web accessibility into account, a sense of responsibility to people with disabilities appears to provide the greatest incentive. Similarly, while many factors prevent web developers from creating accessible websites, time and budgetary constraints appear to be the greatest hindrance.

5.4.3.3. Web accessibility testing

Participants were asked whether they ever tested the accessibility of their websites with disabled and/or older users. In the contextual inquiry study, none of the participants described conducting accessibility testing with users. Only three of the 26 participants

(12%) in this study claim to do so (and two of them claim to do so only occasionally). This is consistent with their responses to the previous mental models questions, which suggested a sparse and haphazard awareness of web accessibility evaluation methods.

One participant (4%) explains that his company conducts "UX sessions" for which they recruit different types of users. Although some of these users may be disabled, he admits that they do not purposely recruit people with disabilities. The other two participants both state they would like to conduct user testing in every project they work on, particularly as they feel the outcome is often more revealing than other evaluation methods. They both stress however, that user testing only occurs when clients and web commissioners were willing to pay for it.

The remaining 23 participants (88%) never conduct accessibility testing with users. They cite a variety of obstacles preventing them do so, the most common being that they have neither the time nor the budget either to conduct the testing or to react to the considerable amount of feedback it generates. Two participants (8%) claim they do not have access to disabled users or that it is difficult to find "typical" disabled users. Another two participants (8%) claim the websites they develop require too much domain-specific knowledge and that disabled users are not part of their target audience. One participant (4%) claims that user testing is the responsibility of someone else, such as a web designer. Another participant (4%) claims that instead of conducting user testing he would prefer to react to feedback from disabled users if and when they encounter a problem.

The time and budgetary pressures that prevent web developers from creating accessible websites also inevitably impact their ability to test websites with users. Unfortunately, as well as precluding awareness of genuine accessibility problems in websites, this myopic approach prevents web developers from gaining insight into how people actually use websites and the accessibility barriers they may encounter.

5.4.3.4. Client expectations

Participants were asked whether the people who commission their websites require or expect them to make their websites accessible. Unsurprisingly, only three of the 26 participants (12%) claim that clients specifically require them to make their websites accessible. Of these, only two participants (8%) claim their clients require them to meet a particular level of accessibility, in both cases the requirement is WCAG Level AA. Five participants (19%) describe how clients occasionally pay lip service to web accessibility, making either vague requests such as "ensure the website is accessible" or very specific requests (such as ensure the website conforms to WCAG level AAA) without any real awareness of the work involved to achieve it. Five participants (19 %) have no direct interaction with the people who commission their websites.

Participants who interact directly with clients and website commissioners were asked whether they were required or expected to *test* the accessibility of their websites. Only one participant (4%) has dealt with a client requesting accessibility testing but it was sub-contracted to an external accessibility consultancy. Two participants (8%) have dealt with clients who have occasionally requested accessibility testing but only in response to specific complaints. Many participants stress that they are not expected to develop accessible websites, let alone test the accessibility of them.

Web developers are rarely required or expected to create or test accessible websites, either by clients or the organisations for whom they work. This lack of buy-in and support from clients and website commissioners makes it very difficult for them to find the time or budget to make websites accessible. Consequently, they are forced either to introduce web accessibility covertly under the guise of 'good practice' or to develop a solid business case for its inclusion.

5.4.3.5. Web accessibility resources

The 22 participants (85%) who currently attempt to make their websites accessible were asked to describe the resources they use. Of these, only 8 participants (36%) claim to use WCAG. Three participants (14%) claim to use other W3C specification documents, including HTML and WAI-ARIA. Two participants (9%) claim to use product-specific accessibility guidelines (e.g. Adobe Flex or jQuery). Three participants (14%) claim to use to more general, informal accessibility guidelines, such as the RNIB 'See it right' documentation or W3Schools tutorials. Four participants (18%) claim to rely upon Google search results to find relevant blogs, articles and tutorials. Only one participant (5%) claims to rely upon print books to assist in making websites accessible.

Instead of using specific accessibility resources, some participants rely on the feedback from web accessibility evaluation tools. These include code validators (18%, 4); browser toolbars (5%, 1), the accessibility features of integrated development environments (IDEs) (5%, 1); and specific accessibility utilities, such as colour contrast analysers (5%, 1). One participant (5%) also relies upon the feedback from ATs, such as JAWS screen reader. Two participants (9%) rely only upon acquired knowledge or 'rules of thumb', whereas four participants (18%) claim not to rely upon any accessibility resources at all.

As in the contextual inquiry study, web developers haphazardly approach the practical aspects of web accessibility evaluation. While some draw on a disparate array of accessibility resources and tools, others rely more on intuition and half-remembered rules of thumb to determine the accessibility of websites. This unsystematic approach may stem from the participants' limited mental models of web accessibility. Without a clear understanding of the problems that disabled people encounter in using the web, web developers are unlikely to determine appropriate solutions.

5.4.3.6. Web Content Accessibility Guidelines

Although participants referred to a variety of resources for making websites accessible, their familiarity and experience of using WCAG was of particular interest, given its prominence and role as an international standard of web accessibility.

Participants were asked whether they were familiar with WCAG and whether they had ever tried to use it. Over a third (38%, 10) claim to be familiar with WCAG, with some noting that WCAG conformance is a client requirement. Half of the participants (50%, 13) claim to be aware that WCAG exists but are not very familiar with it. Only three participants (12%) claim to be completely unfamiliar with WCAG.

Despite this reasonable awareness, few use WCAG on a regular basis. Seven of the 26 participants (27%) claim to have never used WCAG. Ten participants (38%) claim to have used WCAG recently, whereas nine participants (35%) claim to have used WCAG in the past but have not done so recently. Several participants used WCAG earlier in their career (perhaps as part of their training or to meet the requirements of a particular client) but have not used it since. Consequently, these participants appear more familiar with WCAG 1.0 than WCAG 2.0.

While awareness of WCAG is reasonably high among web developers, actual usage is low. Furthermore, participants appear to be more familiar with the initial version of the resource than its more recent incarnation.

5.4.3.7. WCAG conformance levels

Participants were asked whether they were familiar with the three conformance levels of WCAG (A, AA, and AAA) and whether they had ever tested websites against them. Just over a quarter of the 26 participants (27%, 7) claim to be very familiar with the three conformance levels of WCAG (A, AA, and AAA). Some participants claim to be familiar with only one or two levels (e.g. up to Level AA) and some were clearly referring to the numeric Priority Levels of WCAG 1.0. Half of the participants (50%) claim to be aware that the conformance levels exist but are not very familiar with them and cannot provide details of the specific success criteria within each conformance level. Six participants (23%) claim to be completely unfamiliar with the conformance levels.

Just over a quarter of the 26 participants (27%, 7) claim to test to the WCAG conformance levels. Three participants (12%) claim to test to some or all of the conformance levels but not as part of a formal process. Some claim to find the conformance levels useful for filtering and constraining WCAG. Others claim to find the levels useful for benchmarking progress of a website's accessibility. Almost two thirds of participants (62%, 16) claim not to test to any of the conformance levels.

The most common reason for conducting WCAG conformance testing, mentioned by six participants (23%), was if a client requires a particular level of conformance. Two participants (8%) claim WCAG conformance testing is part of their company's QA testing procedure. One participant (4%) claims to conduct conformance testing more in the interest of creating standards conformant websites than achieving accessibility. Only one participant (4%) claims to conduct conformance testing simply as a matter of course.

Although web developers are generally familiar with the concept of WCAG conformance levels, few are aware of what they entail. Often web developers are only exposed to the conformance levels via a web accessibility evaluation tool. For those that are familiar with them, the conformance levels provide a useful benchmark against which to monitor a website's accessibility and also appear to make the resource more manageable and easier to use.

5.4.3.8. Ease of using WCAG

The 19 (73%) participants who had ever tried to use WCAG were asked how easy or difficult they found it to use. Of these, five participants (26%) claim they find WCAG easy to use. They consider WCAG to be "simple" or "straight-forward" to use and implement. One participant (5%) finds the information in WCAG to be "good quality". Another participant (5%) finds WCAG "really interesting". One participant (5%) feels that WCAG supports a "common-sense, best practice approach" and does not contradict what he wants to do with websites. Notably, the participants who are very positive about WCAG are the ones who have used it in the past but have not done so recently and appear to be referring to the first version of the guidelines.

Eleven participants (58%) claim they find WCAG difficult to use. One participant (5%) believes WCAG has become "increasingly difficult" to use and that WCAG 1.0 was much easier to follow than WCAG 2.0. Consistent with the contextual inquiry study, the difficulties appear to arise from several aspects of the guidelines. Some participants find the organisation and structure of WCAG difficult to navigate and understand. One participant (5%) feels there is no clear, high-level description of the content. Some participants find the sheer volume of content in WCAG "overwhelming", "too detailed" and "time-consuming". Several participants also find the language and content of WCAG difficult to understand and feel it is badly written. Other participants feel the guidance provided in WCAG is "too qualitative", "too abstract", "vague and ambiguous" with "nothing very well defined". One participant (5%) finds WCAG "difficult to read, understand and take in" and considers it particularly ironic that accessibility guidelines are themselves inaccessible. One participant (5%) considers certain aspects of WCAG irrelevant whereas another participant (5%) struggles to relate the guidance to the disabled people it is intended to benefit.

The difficulties that web developers encounter in using WCAG are consistent with the contextual inquiry study, and relate to the organisation and structure, the volume and comprehensiveness, and the language and terminology of the guidelines. Without access to an appropriate amount of clear, concise and precise accessibility information that they can easily interrogate, understand, and apply to their work, web developers' mental models of web accessibility are unlikely to develop the comprehensiveness and sophistication necessary to adequately support people with disabilities.

5.5. Discussion

In this study, 26 professional web developers were interviewed with the aim of eliciting their mental models of web accessibility. Drawing upon a number of direct and indirect mental model elicitation techniques from the literature, the study explored what web developers understand about web accessibility and looked to identify any knowledge gaps, inconsistencies and uncertainties that may result in inaccessible websites. It also sought to uncover any misconceptions or faulty thinking that informs web developers' current approaches to web accessibility. Ultimately, by understanding web developers' mental models of web accessibility, the study sought to gain greater insight into how they currently approach their work.

In describing what they understood by the term "web accessibility", each participant provided a reasonably accurate and consistent definition of the term, characterising it as a process of developing websites that are usable and accessible to all people, including those with disabilities. Their common understanding of web accessibility is encouraging but unsurprising. It is consistent with the established definition of the term common to many guidelines, textbooks, blogs, and other information resources. For instance, according to the WAI: "Web accessibility means that people with disabilities can use the Web" (Henry, 2005a, para. 1). Similarly, the introduction to WCAG 2.0 describes how the document "defines how to make Web content more accessible to people with disabilities" (Caldwell et al., 2008a, para. 1). The WebAIM Introduction to Web Accessibility document aims to "help you understand how people with disabilities use the web" (WebAIM, 2014, para. 2). Wikipedia also defines web accessibility as the "practice of removing barriers that prevent access to websites by people with disabilities" ("Web accessibility", n.d., para. 1). It is reasonable to assume that even web developers who make little attempt to create accessible websites are likely to have encountered the term before, resulting in at least a basic understanding of its meaning.

In spite of participants' apparent user-centred understanding of web accessibility, their awareness of the beneficiaries of accessible websites is limited. People with visual disabilities and people with physical disabilities dominate the participants' mental models of web accessibility, with other user groups, such as deaf and hard of hearing people, people with specific learning disabilities, and older adults, featuring much less prominently. This limited awareness of disabled people also reflects in their knowledge of ATs that people use. The majority of participants seem to understand that people with visual disabilities either cannot or have difficulty seeing screens and typically use screen readers to read aloud the contents of webpages. Many participants are also aware that some physically disabled people lack the finger or hand dexterity to use a regular keyboard or mouse and rely on alternative input devices to interact with websites. The participants' awareness of other ATs, however, is limited. With only a limited awareness of the beneficiaries of web accessibility and the ATs they use, participants' understanding of how to create and evaluate accessible websites is also impeded. Despite referring to a range of accessible coding practices, participants appear most familiar with those that make websites easier to use with screen readers and keyboards. Similarly, while participants generally recognise the need to evaluate the accessibility of websites, their practical knowledge is sparse and haphazard.

Overall, participants' mental models appear to incorporate some, *but not nearly enough*, knowledge and awareness of web accessibility. Their knowledge of how to create and evaluate accessible websites is predicated on a very limited understanding of disabled people, the ATs they use, and the challenges they face in using the web. Consequently, the extent of their support is limited to employing specific coding practices (e.g. providing text alternatives or making font and colour adjustments) to support specific user groups (e.g. people with visual or physical disabilities) using specific ATs (e.g. screen readers and alternative input devices) with no specific awareness of how to evaluate the outcome.

Despite the apparent limitations of participants' web accessibility knowledge, the accuracy of their predictions suggests they can adequately "run" their mental models to predict the consequences of particular actions. The participants' tasks (in this case, supporting different user groups) appear to be adequately mapped or yoked to the appropriate actions (coding techniques that support those user groups). These so-called task-action mappings guide participants in the selection of appropriate coding techniques to benefit particular user groups and vice versa. The fact that participants are able to do this even for less familiar stimuli is encouraging. It suggests that they have the cognitive infrastructure to develop websites in an accessible way. Young (1981) considers task-action mappings to be particularly effective in allowing people to generalise from familiar to novel tasks. Once people have established a core set of task-action mappings, Young argues, they can easily generalise this knowledge to perform novel tasks.

Alternatively, participants may have based their predictions on a simple means-end analysis. This was first proposed by Newell and Simon (1972) as an effective but inefficient problem solving heuristic. By envisioning the end or ultimate goal, the problem solver determines the best sequence of actions from their current position that will lead to the goal. Thus, given the goal of supporting a particular user group (e.g. partially sighted people), participants may simply match the characteristics of that group (e.g. low vision) to the technologies they may use (e.g. screen magnifiers), and then match those technologies to the appropriate coding techniques (e.g. allow content can be resized). Certainly, given the apparent limitations of their web accessibility knowledge, a means-end analysis seems a more likely approach for participants to have taken on this occasion.

Underpinning the participants' procedural knowledge of web accessibility is a conceptual model that appears to prioritise technical compatibility over user needs. Participants have only a superficial understanding of users and perceive them not in terms of how they use websites but in terms of the technologies they use in doing so. Many participants describe making web content "available" or "friendly" not to disabled people, but to screen readers and other ATs. Some participants even refer to disabled people by the technologies they use, such as "screenreader users", "keyboard users", and "non-mouse users". Similarly, participants frequently refer to accessible coding practices in terms of how they facilitate ATs, devices and user agents to access the web, with little apparent awareness of how they ultimately benefit disabled users.

Participants' apparent prioritisation of technology over users may represent a generalisation of their existing mental models of web development. It is widely accepted in mental models theory (e.g. Young, 1981; Norman, 1983) that people apply, with varying success, their existing mental models of familiar situations to unfamiliar (but seemingly related) situations. Web developers, who are already adept at ensuring their websites are compatible with a range of browsers and devices and who are likely to have a mental model that reflects this, may be applying the same mental model to *accessible* web development. Making websites accessible, therefore, becomes just another layer of compatibility to be achieved. For web developers with technology-oriented mental models, a screen reader or an alternative input device represents just another user agent with which they need to ensure their websites are compatible.

The conceptual model underpinning participants' mental models of web accessibility also appears to prioritise standards conformance over user experience. This results in participants perceiving web accessibility as a technical, rather than human, challenge. In defining web accessibility, many participants draw upon terms such as "making sure" and "ensuring" that characterise it as a process of conformance as opposed to a more creative endeavour. Some participants describe making webpages accessible through the use of semantic and standards-compliant HTML. They also evaluate the accessibility of websites by determining whether they conform to published standards and guidelines or render correctly across different user agents and devices. Rarely does their approach involve or consider actual users. However, as studies by Freire (2012), the DRC (2004), and Rømen and Svanæs (2008, 2011) have demonstrated, while technical compliance and standards conformance may go some way to ensuring a website can be *accessed* by people with disabilities, it does not guarantee that websites will be *accessible*.

Existing tools, guidelines and resources that present web accessibility as a technical challenge may have reinforced web developers' conformance-oriented mental models. For example, automated accessibility evaluation tools tend to take the form of code validators. Such tools inspect website code and present a checklist of issues to address, often with no rationale or connection to the users who will benefit. Similarly, the role of WCAG as a technical standard of web accessibility with varying levels of conformance further promotes the idea that web accessibility is simply a matter of technical compliance. Sloan and Kelly (2008) argue that WCAG's role as a technical standard appeals to web developers' desire for technical quality of their work but does little to raise awareness and understanding of users. For many web developers, they argue, web accessibility is defined by standards-conformance as opposed to usability by humans. Without a more nuanced understanding of disabled people and an awareness of the techniques and strategies they use, the support that web developers can provide is limited to merely ensuring websites are technically compliant with standards and guidelines. Web developers are constrained to creating websites that can be *accessed*, but not necessarily *used*, by people with disabilities.

The participants' responses to questions about their working practices and attitudes to web accessibility were consistent with the findings of the contextual inquiry study and provide context to their other responses. For instance, the majority of participants claim to be motivated by professional pride or a desire to "do the right thing" by creating websites that are accessible by as many people as possible. However, these aspirations are often dashed by client demands and organisational attitudes that place little value on accessible websites. This suggests that the limitations of participants' mental models stem not from lack of awareness or concern but from organisational constraints on their time. These constraints prohibit participants from investing sufficient time in learning about and implementing web accessibility. This is not helped by a lack of adequate knowledge and practical guidance on how to make websites accessible. Participants familiar with WCAG encounter many of the same problems as their counterparts in the contextual inquiry study. These include difficulties relating to the language and content, the organisation and structure, and the size and volume of the guidelines. Given WCAG's role as an international standard of web accessibility, it is inevitable that participants' experiences of the resource will have contributed to their perception and understanding of web accessibility. Although the contextual inquiries generated a considerable amount of rich, meaningful data about the working practices of web developers, the study sample was limited to only 13 participants. Therefore, the fact that consistent findings emerged from a different and much larger sample of web developers suggests these findings accurately reflect the experiences this population.

The limitations of this study largely stem from the ambiguous and intangible nature of mental models. In lieu of a robust definition of the concept, it is impossible to elicit or represent someone's mental model with any certainty. This is particularly the case for such a broad concept as web accessibility. Unlike mechanical devices or interactive systems, for which one can establish at least a (correct or incorrect) basic working knowledge, web accessibility represents a much broader, more abstract practice. Consequently, it is difficult to determine whether someone actually has a mental model of web accessibility, let alone to be able to elicit it.

To overcome this uncertainly, I combined a number of direct and indirect elicitation techniques from the literature, none of which were without their limitations. For example, asking participants to describe what they understood by the term "web accessibility" may have generated an appropriate and, in many cases, detailed response but may not have truly elicited their mental model of the concept, only whatever came to mind at the time. Similarly, asking participants to predict the outcome of different actions have merely tested their reasoning abilities rather than eliciting their mental models. Increasing the difficulty of this exercise may have more accurately reflected their understanding of web accessibility. Finally, while interviewing participants about their existing working practices and attitudes towards web accessibility served to reaffirm many results from the contextual inquiry study with a different and larger population, participants may have felt obligated to provide more socially-acceptable responses, particularly in regard to their attitudes towards web accessibility. Nevertheless, the combination of techniques used in this study gave further insight into the "developer mindset" and highlighted a number of issues to consider regarding the provision of web accessibility support to web developers.

5.6. Conclusions

The findings of this study and the previous study present a complex challenge for the development of accessibility tools, guidelines and resources. On the one hand, such resources need to address the misconceptions, correct the errors and fill in the knowledge gaps of web developers in order to further develop their understanding of web accessibility. They also need to shift the "developer mindset" from satisfying a collection of technical requirements, to meeting the needs of as many users as possible. Such an approach should make web accessibility less of an onerous, seemingly arbitrary tick-box exercise and turn it into an efficient, fulfilling activity. However, web developers must still operate under certain organisational constraints that may be out of their control. Therefore, in addition to presenting a more user-centred understanding of web accessibility, accessibility tools, guidelines and resources must respect web developers' existing working practices. They need to provide sufficient detail so that web developers have a good understanding of what they are doing but not so much detail that the practical techniques are obscured, and their workflows are disrupted.

In essence, accessibility information resources need to promote a more humanistic portrayal of web accessibility that focuses less on developing technically compliant code and more on supporting the needs and requirements of disabled users. For this to succeed, however, it needs to respect web developers' existing working practices. This study highlights the importance of a strong foundation of knowledge on which to base accessibility guidance. For web developers to be effective in employing web accessibility, they need to have an understanding of who it benefits and how. However, the information must be presented in a way that is familiar and understandable to web developers and easily applicable to their work.

Chapter 6. Design, implementation and initial evaluation of a web accessibility information resource

6.1. Introduction

The previous chapters described the analysis stage of this programme of research, in which I drew partly upon ethnographic methods to explore the current working practices of professional web developers and to examine the current state of web accessibility. These chapters reaffirmed many of the financial, political and organisational constraints that prevent web developers from incorporating accessibility into their day-to-day web development activities. Web developers are rarely afforded the time and resources necessary to make websites accessible. They face considerable resistance among other stakeholders (in particular, management), who continue to perpetuate misconceptions about the cost and challenges of web accessibility. Accessible websites are frequently misperceived as expensive, aesthetically displeasing and over-simplified substitutes that only benefit a small proportion of the population.

There are evidently still significant barriers to web accessibility becoming commonplace. While fully overcoming these barriers would require substantial organisational or industrial change, a more achievable solution may lie with web developers themselves. The previous chapters demonstrate how web developers' limited mental models of web accessibility bear little theoretical understanding of disability issues or practical knowledge of how best to provide accessibility support. This motivates a clear need for practical, user-centred accessibility guidance to web developers that furthers their knowledge and understanding of web accessibility while respecting their working practices. Existing tools, guidelines and resources let web developers down by failing to provide them with the support and information they need, in the way that they need it. This thesis specifically addresses this aspect of the problem, focusing on the creation of resources to educate and inform web developers.

6.1.1. Existing accessibility information resources

As described in Section 2.2, numerous web accessibility initiatives over the last two decades have resulted in a well-established body of accessibility information, often presented in the form of a set of guidelines or recommendations, such as WCAG and Section 508. Recognising the challenges of using resources such as WCAG in their raw, unabridged state (see Section 2.5.3), researchers and industry experts have explored alternative means of delivering accessibility support and information to web developers in a more meaningful and useful way. Such efforts have largely focused on managing, administering and clarifying the existing sets of guidelines.

For example, the MAGENTA tool (Leporini, Paternò, and Scorcia, 2006) assists web developers in defining, handling and checking multiple sets of accessibility guidelines (including WCAG; the Italian Stanca Act that applies to Italian government websites; as well as custom guidelines developed by the authors for visually impaired users). It performs semi-automatic accessibility evaluations and provides web developers with advice on how to address accessibility issues. While the MAGENTA tool has the potential to improve the management and application of accessibility guidelines, particularly where multiple sets of guidelines are applicable, it does little to clarify or contextualise them. Furthermore, the semi-automatic nature of the tool constrains its advice to programmatically determinable issues.

Recognising flaws in existing accessibility guidelines, Donnelly and Magennis (2002) developed a new set of Irish national IT accessibility guidelines, engaging various stakeholders (including web developers) in a user-centred development process. They highlighted numerous aspects that may have prevented the wider adoption of existing guidelines (including WCAG 1.0), such as language ambiguities, the considerable level of technical knowledge of accessibility necessary to read and interpret the guidelines, and lack of motivation to make websites accessible. Their proposed guidelines provide a more usable interface to existing guidelines by expanding upon and clarifying them but still maintain their existing organisation and structure.

The Accessibility Evaluation Assistant (AEA) (Bailey and Pearson, 2010) is an educational tool to support novice accessibility evaluators in conducting accessibility evaluations. Similar to the guidelines developed by Donnelly and Magennis, the AEA presents a series of tailored checkpoints by filtering web accessibility guidelines, such as WCAG, according to particular user groups or types of web content. In addition to streamlining the evaluation process, the filtering of guidelines is designed to familiarise novice evaluators with different user groups and contexts of use. Although this approach could potentially benefit web developers, the AEA is intended primarily for students to develop skills in web accessibility evaluation. Despite the AEA being well-received by this population, an evaluation of the tool with novice and expert evaluators (Bailey and Pearson, 2011) found low levels of agreement both among the novices and between the two groups.

WebAIM (Web Accessibility In Mind)[52] is a non-profit organisation within the Center for Persons with Disabilities at Utah State University. Established in 1999, WebAIM offers a comprehensive range of services, tools and resources about web accessibility that are popular with web developers. These include a collection of articles that attempt to clarify and explain the more esoteric aspects of web accessibility, including the various guidelines, laws and standards. Each article provides an in-depth user-centred exploration of a particular topic (e.g. semantic structure; alternative text; keyboard accessibility). While this inevitably includes accessible techniques, solutions and examples, the long-form nature of the articles encourages a more considered read. Consequently, WebAIM's articles are better suited to someone looking to further their knowledge of the subject than someone seeking practical guidance on web accessibility evaluation. WebAIM does provide a distillation of its practical guidance in a *WCAG 2.0 Checklist*[53], but, as with the solutions discussed above, this merely repackages and rephrases the existing guidelines while maintaining their organisation and structure.

As described in Section 2.2, the WAI's EOWG also provide a range of support materials aimed at making the existing information in WCAG more digestible and comprehensible to web developers (Henry, 2017b). For example, *Easy Checks*[54] offers guidance on conducting a first review of web accessibility. Despite its connection to WCAG 2.0, the resource eschews the more complex language and terminology of its parent resource in favour of clear, user-centred practical information. However, due to being aimed very much at beginners, *Easy Checks* covers only a limited selection of easily identifiable accessibility issues (page titles; image text alternatives; headings etc.). A more recent WAI initiative is *Web Accessibility Tutorials*[55], which provide more detailed guidance on how to create WCAG-conformant websites. The tutorials cover a (currently limited) selection of accessibility topics (page structure, menus, images etc.)

---

[52] https://webaim.org/
[53] https://webaim.org/standards/wcag/checklist
[54] https://www.w3.org/WAI/eval/preliminary
[55] https://www.w3.org/WAI/tutorials/

and, much like WebAIM's articles, their in-depth long-form nature necessitates a much greater commitment.

Despite a common objective in improving access to and comprehension of accessibility information, none of these efforts appear to have been designed specifically to accommodate web developers' existing working practices and organisational constraints or to exploit their existing knowledge and understanding of web accessibility.

6.1.2. WebAIR: Web Accessibility Information Resource

To address the various problems that hinder access to web support and accessibility information by web developers, and to embody the research findings of the previous stage of this programme of research, I designed a prototype accessibility information resource, called WebAIR: Web Accessibility Information Resource. The purpose of WebAIR—whose design, implementation and evaluation are described in this chapter—is both to support web developers in the creation and evaluation of accessible websites and to foster a greater understanding of web accessibility and its beneficiaries.

It is important to stress that the intention behind WebAIR is not to supplant existing accessibility guidelines, such as WCAG 2.0 (which, in many countries, is enshrined in law), but to complement them, by providing supplementary guidance and clarification that is tailored specifically to the needs and requirements of web developers. WebAIR attempts to convey the principles and concepts of web accessibility established in WCAG 2.0 but in a more usable and relevant format. It goes beyond WCAG 2.0 to deliver the timely, practical guidance that web developers need to do their job effectively, while also furthering their knowledge and promoting a user-centred conceptual understanding of web accessibility.

It should be noted that WebAIR was initially created under the auspices of the i2Web project, where it was known as the *i2Web Content Taxonomy*. The initial version of the resource was created in response to the outcome of the contextual inquiry investigation presented in Chapter 4 and integrated into an accessible web development tool, called the Evaluation of Accessibility Support and Integration (EASI) Tool that had been developed as part of the i2Web project. The EASI Tool was deployed as a plug-in for the Eclipse IDE[56], a popular open-source development environment. It integrated the results of automated and semi-automated testing tools to assist web developers in undertaking both formative and summative accessibility evaluations. These evaluations were supported by material drawn from WebAIR to provide relevant and timely accessibility guidance (see Section 6.5.1.3 for a more detailed account of the tool's interface). Despite the integration of WebAIR material, the EASI Tool was primarily intended as an automated accessibility evaluation tool, with little emphasis placed on educating and informing web developers. As observed in Chapter 4, the use of such tools contributes little to web developers' understanding of web accessibility and may even promote a false sense of confidence. Therefore, while tool integration might be a beneficial approach for future development (see Section 10.5), the focus of this thesis is on the design and evaluation of WebAIR as a standalone resource.

WebAIR includes a considerable amount of original support material this is written in web development-oriented language and organised around the working practices of web developers. Although I conceived the resource and defined its initial structure and organisation, several other members of the i2Web project consortium contributed content. Consequently, the initial prototype of WebAIR presented in this chapter is inconsistent in tone and quality. Furthermore, as only a proof of concept was necessary for the i2Web project, some sections of the resource are incomplete. This limitation was addressed in the subsequent version of WebAIR, which, in response to the outcome of the initial evaluation described in this chapter, I substantially revised and rewrote. As well as developing a more consistent tone and quality, the revision allowed me to incorporate the results of the study of web developers' mental models of web accessibility presented in Chapter 5, which I conducted in parallel to this initial evaluation using the same group of participants. An abridged version of the development and initial evaluation of the resource was included in the i2Web project deliverables, *D4.7: Final user-based evaluations of accessibility and usability of services and applications* (Petrie, 2013a) and *D4.8: Field studies of the use of i2Web services and applications* (Petrie, 2013b). The revision of WebAIR and its subsequent evaluations were not part of the i2Web project; I exclusively conducted them as part of my PhD investigation.

This chapter describes the design, implementation and initial evaluation of WebAIR. A primary evaluation was conducted with 26 professional web developers, who had the

[56] https://www.eclipse.org/ide/

opportunity to try out complete and abridged versions of both WebAIR and WCAG 2.0 in an ecologically valid setting. A smaller, secondary evaluation of WebAIR was conducted with 7 computer science students, who may be classed as novice web developers. The novices provided feedback on the version of WebAIR that had been integrated into the i2Web EASI Tool. The chapter concludes by describing how WebAIR was revised in response to this initial evaluation.

Despite one of the aims of WebAIR being to further web developers' knowledge and understanding of web accessibility, this aspect of the resource could not adequately be evaluated in this study. Limitations on time meant that participants in both evaluations were only briefly exposed to each resource, making any observable changes in knowledge or understanding unlikely. Instead, the study objective was to gain the web developers' initial impressions of the different resources and to measure the impact of different design decisions made in the development of WebAIR on web developers' attitudes and approach to web accessibility.

6.2. Design and implementation of WebAIR

The problem statement for this stage of the research was to design an accessibility information resource to support web developers to create and evaluate accessible websites. Crucially, the resource should not only promote the procedural or technical aspects of accessible web development but also stimulate a broader conceptual or theoretical understanding of web accessibility and its beneficiaries. This should be achieved without unduly disrupting established workflows. The resource should take the form of a standalone, multi-page website that could potentially be integrated into existing tools and testing frameworks.

To tackle this problem statement, several key design decisions were made. Each decision addressed one or more requirements that had emerged from the previous studies concerning the difficulties that web developers face in creating and maintaining accessible websites (see Table 6.1). The subsequent sections describe the realisation of these requirements in the design and implementation of WebAIR.

Table 6.1: Requirements for a web accessibility information resource

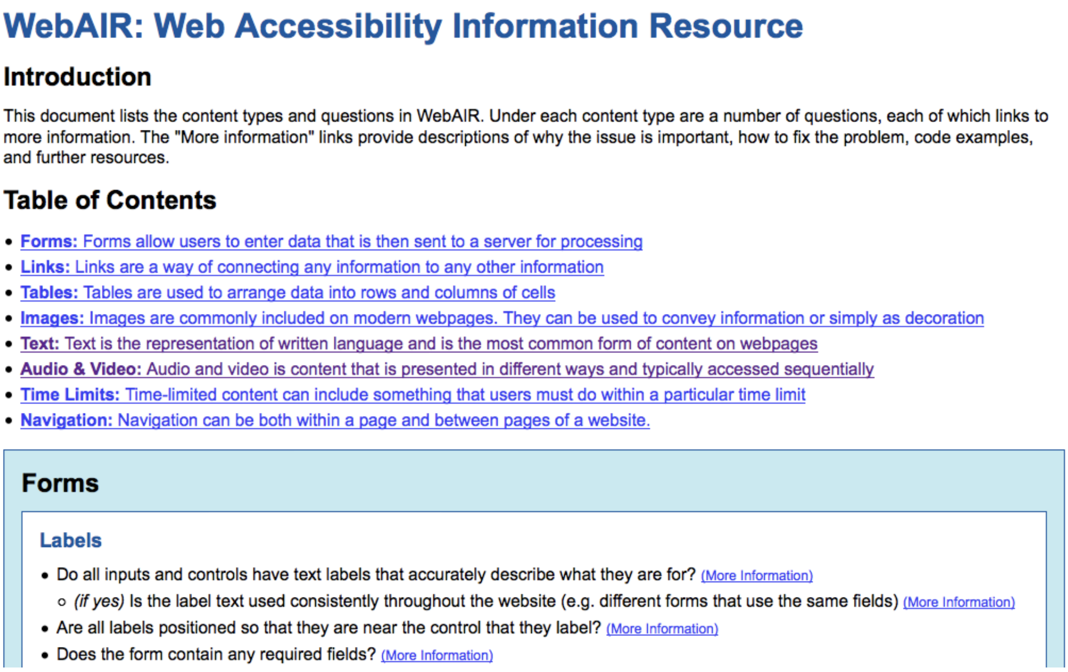
ID	*Requirement*	*Rationale*	*Source*
RQ1	Must provide specific, unambiguous guidance to web developers.[57]	Existing tools, guidelines and resources tend to be couched in vague and indefinite statements.	Section 4.3.5.3; 5.4.3.8
RQ2	Must use language and terminology that is familiar to web developers.	Existing tools, guidelines and resources assume web developers are familiar with the domain-specific concepts of web accessibility.	4.3.5.3; 5.4.3.8
RQ3	Must provide a justification or rationale behind the guidance to educate and inform web developers.	Web developers struggle to relate existing accessibility guidance either to their work or to the people it is intended to benefit.	4.3.5.2; 5.4.3.8
RQ4	Must be organised according to web developers' content-oriented workflows.	Existing tools, guidelines and resources tend to be organised according to web accessibility concepts that web developers find difficult to navigate and apply to their work.	4.3.5.2; 5.4.3.8
RQ5	Must allow web developers to dip into the guidance as appropriate.	Web accessibility is typically left to the very end of development, instead of being integrated throughout the web development process.	4.3.4.1;
RQ6	Should promote user experience over standards conformance.	Existing tools, guidelines and resources tend to present web accessibility as a technical, rather than human challenge.	4.3.5.2; 5.5
RQ7	Should encapsulate coverage of WCAG 2.0.	Web developers are often required to ensure their websites conform to a specific accessibility standard, such as WCAG 2.0.	4.3.2.4; 5.4.3.4
RQ8	Must provide a reasonable and manageable amount of information.	Existing tools, guidelines and resources tend to include an overwhelming amount of information that web developers find verbose.	4.3.5.1; 5.4.3.8
RQ9	Must provide a realistic number of items for web developers to check.	Existing tools, guidelines and resources tend to include a considerable number of items for web developers to check that they find time-consuming.	4.3.5.1; 5.4.3.8

[57] WebAIR is designed to specifically support web developers. However, as noted in Section 1.3, the responsibilities of web developers can vary from one organisation to another, making the role difficult to precisely define. Some organisations might, for example, consider copywriting to be the responsibility of web developers, whereas others may employ a dedicated copywriter. Despite the potential for some of WebAIR's guidance to be rendered irrelevant to some, the resource was written with the broadest definition of web developer in mind.

Of course, the focus on web developers should not preclude people in other roles (such as copywriters) from benefiting from the guidance in WebAIR. As proposed in Section 10.5, the resource could, in future, potentially be adapted for other roles in the value chain of accessible web development to allow web accessibility to be integrated into all stages of design and development.

6.2.1. Language and terminology

WebAIR comprises a set of specific, unambiguous questions that address different accessibility issues **[RQ1]**. Each question represents a call to action that asks web developers whether they have completed a particular task (see Figure 6.1). For example, to determine whether websites are keyboard accessible, it asks web developers: *Can you successfully access all links using the keyboard?* To address any ambiguities around form submission, it asks web developers: *Do you provide feedback when a form has been submitted successfully?* The questions are constructed to avoid the domain-specific language of web accessibility commonly found in existing tools, guidelines and resources **[RQ2]**. Instead, the questions incorporate web development terms or refer to specific user actions in the interface.



WebAIR: Web Accessibility Information Resource

Introduction

This document lists the content types and questions in WebAIR. Under each content type are a number of questions, each of which links to more information. The "More information" links provide descriptions of why the issue is important, how to fix the problem, code examples, and further resources.

Table of Contents

- **Forms:** Forms allow users to enter data that is then sent to a server for processing
- **Links:** Links are a way of connecting any information to any other information
- **Tables:** Tables are used to arrange data into rows and columns of cells
- **Images:** Images are commonly included on modern webpages. They can be used to convey information or simply as decoration
- **Text:** Text is the representation of written language and is the most common form of content on webpages
- **Audio & Video:** Audio and video is content that is presented in different ways and typically accessed sequentially
- **Time Limits:** Time-limited content can include something that users must do within a particular time limit
- **Navigation:** Navigation can be both within a page and between pages of a website.

Forms

Labels

- Do all inputs and controls have text labels that accurately describe what they are for? (More Information)
  - *(if yes)* Is the label text used consistently throughout the website (e.g. different forms that use the same fields) (More Information)
- Are all labels positioned so that they are near the control that they label? (More Information)
- Does the form contain any required fields? (More Information)

Figure 6.1: Index page of WebAIR

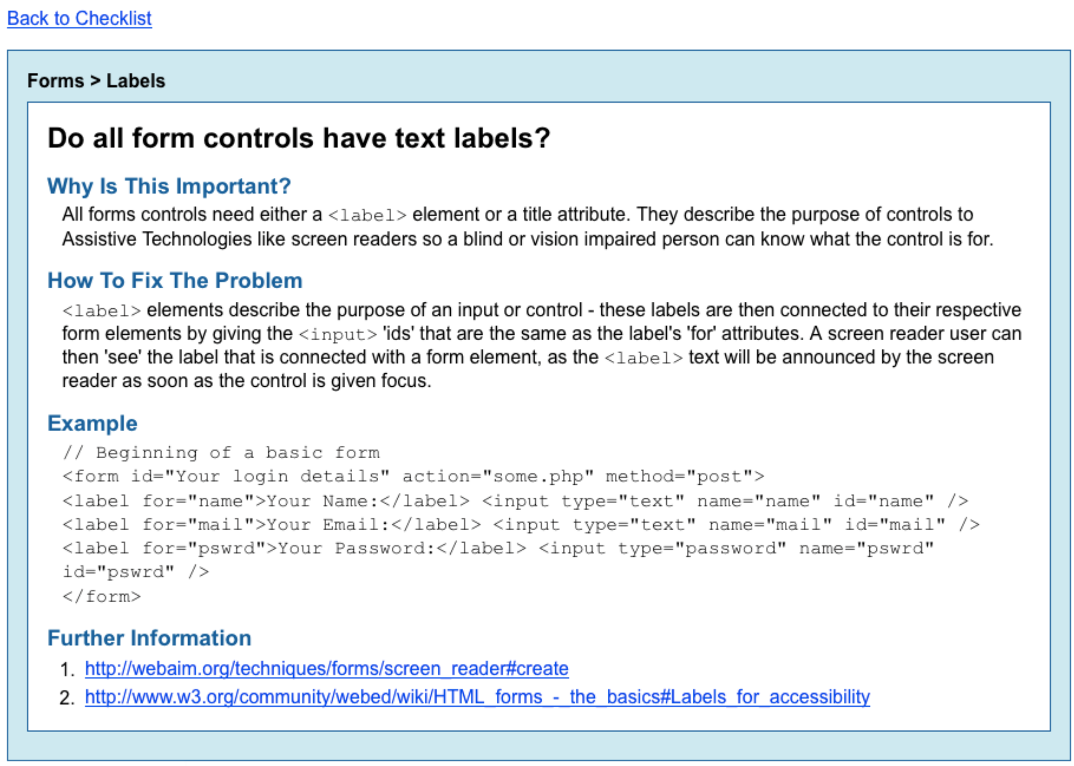
To minimise the initial number of items to check **[RQ9]**, questions that are dependent upon the outcome of other questions are presented as follow-up, subsidiary questions. For example, the question: *Are there any images on the page that flash or blink rapidly?* (which introduces the accessibility issue relating to flashing content, which can be problematic for people with photosensitive epilepsy and other photosensitive seizure disorders) leads to the sub-question: *(if yes) Do they flash less than three times a second?* (which provides more specific guidance). In total, the initial prototype of WebAIR contains 205 questions: 112 first-level questions; 71 second-level questions; and 22 third-level questions.

Each question is accompanied by a 'More Information' webpage (see Figure 6.2), containing only essential information about the accessibility issue it targets **[RQ8]**.

Each 'More Information' page includes:

- a brief (typically one paragraph) rationale as to why that particular question is important and what the consequences of not addressing it would be **[RQ3]**;
- a brief (again, typically one paragraph) explanation of how to address the question and fix the problem;
- a code example, where appropriate; and
- links to further information, where available (these refer web developers to authoritative sources providing more detailed information. For example, the 'More Information' page for the WebAIR question: *Do all form controls have text labels?* includes links to more detailed resources from WebAIM and the W3C).

Unlike resources such as WebAIM, that provide in-depth long-form guidance, which may be difficult to digest quickly, the aim of the 'More Information' pages is to provide web developers with "just-in-time" training in web accessibility principles, allowing them to gradually learn about the domain of web accessibility while they are answering the questions. The advice is presented in a non-prescriptive fashion to accommodate alternative (and potentially superior) approaches to addressing each question.



Back to Checklist

Forms > Labels

Do all form controls have text labels?

Why Is This Important?

All forms controls need either a `<label>` element or a title attribute. They describe the purpose of controls to Assistive Technologies like screen readers so a blind or vision impaired person can know what the control is for.

How To Fix The Problem

`<label>` elements describe the purpose of an input or control - these labels are then connected to their respective form elements by giving the `<input>` 'ids' that are the same as the label's 'for' attributes. A screen reader user can then 'see' the label that is connected with a form element, as the `<label>` text will be announced by the screen reader as soon as the control is given focus.

Example

```
// Beginning of a basic form
<form id="Your login details" action="some.php" method="post">
<label for="name">Your Name:</label> <input type="text" name="name" id="name" />
<label for="mail">Your Email:</label> <input type="text" name="mail" id="mail" />
<label for="pswrd">Your Password:</label> <input type="password" name="pswrd"
id="pswrd" />
</form>
```

Further Information

1. http://webaim.org/techniques/forms/screen_reader#create
2. http://www.w3.org/community/webed/wiki/HTML_forms_-_the_basics#Labels_for_accessibility

Figure 6.2: Typical 'More Information' page from WebAIR

6.2.2. Organisation and structure

WebAIR is organised according to web developers' workflows, specifically their tendency to structure their work according to the types of web content on which they are working, either during web development or evaluation **[RQ4]**. The previously described questions in WebAIR are organised according to different types of web content (e.g. forms, images, tables, text etc.), with the aim of providing web developers with a concrete categorisation that they can more easily apply to the type of content they are working on. For example, a web developer creating or evaluating a form or a table would locate the relevant content category in WebAIR (e.g. *Forms* or *Tables*) and answer each question within.

Each content type comprises several sub-categories to further classify the questions. For example, the 20 questions relating to forms are grouped into the following sub-categories: labelling; grouping; navigating form fields; completing forms; and errors. Overall, the 205 questions are categorised according to 10 content types.

- **Forms**: Allow users to enter data that is then sent to a server for processing (20 questions).
- **Links**: A way of connecting any information to any other information (10).
- **Tables**: Used to arrange data into rows and columns of cells (8).
- **Images**: Commonly included on modern webpages. Can be used to convey information or simply as decoration (6).
- **Text**: The representation of written language and the most common form of content on webpages (26).
- **Multimedia**: Content that is presented in different ways and typically accessed sequentially (15).
- **Interactive content**: The presentation of material, information and experiences which require users to actively choose how they wish to explore content (15).
- **Time-limited content**: Something that users must do within a particular time limit (2).
- **Within pages**: Anything that applies to a webpage on the whole, as opposed to specific content items (6).
- **Between pages**: Anything that applies across a group of webpages (4).

WebAIR is designed with the intention of it being used piecemeal and on demand to provide "just-in-time" training throughout the web development process. To this end, certain questions are repeated across different sections. For example, a question about code validation is repeated in the sections relating to forms, links, tables, text and multimedia content. Similarly, questions relating to text alternatives appear in the sections relating to both images and image links. Although this may appear repetitive when the resource is viewed as a whole, the intention is that web developers will encounter each section in isolation **[RQ5]**.

The order of the different content types (as presented above) is not prescriptive and is not intended to define how the creation or evaluation of accessible websites should be conducted. Having identified in Chapter 5 that web developers appear to have a conformance-oriented mental model of web accessibility that prioritise standards conformance over user experience, the concept of conformance or priority levels was purposely avoided in WebAIR **[RQ6]**. The aim of this approach, in addition to including the justification or rationale behind each question, is to promote user experience over standards conformance.

6.2.3. Volume and comprehensiveness

As established in the previous chapters, application of WCAG 2.0 is often mandatory. Consequently, any resource that attempts to guide the development of accessible websites must at least cover WCAG 2.0 to be acceptable. To achieve this with WebAIR, a mapping was undertaken from existing WCAG 2.0 success criteria and techniques onto each question **[RQ7]**. Each WCAG 2.0 success criterion maps onto one or more questions in WebAIR. For example, WCAG 2.0 success criterion 3.3.2 relates to the provision of labels or instructions when content requires user input. The information contained within this success criterion maps to several questions in WebAIR, including *Do all form controls have text labels?*, *Do you indicate required fields in the label text?*, and *Do you provide an example of how users should complete form fields?* A complete mapping of every WebAIR question to its corresponding WCAG 2.0 success criteria and techniques can be found in Appendix G.

As noted above, to provide a reasonable and manageable amount of information **[RQ8]**, each 'More Information' webpage contains only essential information about the accessibility issue it targets. To provide a realistic number of items for web developers to check **[RQ9]**, the section of each 'More Information' webpage describing how to fix the problem is populated with only one example solution drawn from the numerous techniques provided in WCAG 2.0. While not as comprehensive in the variety of ways in which web accessibility can be achieved, this reduction tackles the information and procedural overload problem that developers currently encounter in attempting to apply web accessibility principles and solutions to their websites.

Another attempt at tackling the problem of information and procedural overload **[RQ8, RQ9]** involves the creation of two versions of WebAIR: a 'complete' version providing full coverage of accessibility issues (mapped from WCAG 2.0), and an 'abridged' version covering only the accessibility problems most frequently encountered by users. The selection of accessibility problems is based upon evidence from a number of studies by the DRC (2004), Petrie, Hamilton, and King (2005), Freire (2012) as well as the results of the user requirements gathering process in the i2Web Project. While not as broad in scope or as comprehensive in the amount of accessibility information provided, the abridged version of WebAIR represents a broader attempt to reduce the amount of information presented to web developers. Evaluation of the two versions will investigate whether the risk of compromising the resource's effectiveness through a reduction in scope is outweighed by a greater enthusiasm for web accessibility from web developers and an increased likelihood of them using the resource.

The following section describes the initial evaluation undertaken to investigate whether the design decisions embodied in WebAIR would improve access to web accessibility information for web developers to better support the creation and evaluation of accessible websites.

6.3. Method

6.3.1. Design

To investigate whether the design decisions embodied in WebAIR improve access to web accessibility information and support the creation and evaluation of accessible websites, a within-participants study was undertaken with professional web developers[58]. The professional web developers used the 'complete' and 'abridged'

[58] The study was conducted alongside data collection for the i2Web Project. I was assisted in the data collection by two researchers working on the i2Web Project: Anna Bramwell-Dicks and Lucy Buykx.

versions of both WebAIR and WCAG 2.0. Their reactions to the resources were elicited using a combination of rating items and semi-structured interview questions.

WCAG 2.0 was chosen as a comparator to reflect the current standard of practice in accessible web development. Given its role as an international standard of web accessibility, WCAG 2.0's influence and importance cannot be ignored. Despite numerous criticisms of the resource, both from the research community (see Section 2.5.3) and from web developers themselves (see Section 5.4.3.8), application of WCAG 2.0 is often mandatory. The preamble to WCAG 2.0 also states that it was designed to meet the needs of, among others, web developers, making it a reasonable standard against which to compare the usability and effectiveness of WebAIR.

The evaluation methodology was piloted on a web developer who is a friend of the author, with the procedure refined in light of that experience.

6.3.2. Participants

Twenty-six professional web developers took part in the study. Three participants were female, the rest were male. The age of participants ranged from 22 to 44 years (mean: 31.0). Three participants were from Italy; the remainder were from various parts of the UK. Five participants worked for i2Web consortium organisations. The others were recruited using opportunistic sampling from web developer mailing lists, online forums, social media, conference flyers and personal contacts in the industry. The participants had between 1 and 14 years of experience of web development (mean: 8.0 years). Six participants worked for large enterprises (250+ employees), fifteen worked for SMEs (< 250 employees) and five were self-employed freelancers.

The study was conducted at the same time as the study presented in Chapter 5. For their participation in both studies, participants were offered £30 worth of Amazon vouchers in compensation for their time and effort.

Anna conducted eight of the twenty-six evaluations, Lucy conducted five evaluations, and I conducted the remaining thirteen evaluations. The analysis is entirely my own.

6.3.3. Materials

The study with professional web developers involved four different types of accessibility information resource.

- The complete set of WCAG 2.0 guidelines (presented to participants using the official WAI documentation[59]).
- The complete version of WebAIR (presented to participants as a standalone multi-page website[60 61]).
- The abridged set of WCAG 2.0 guidelines.
- The abridged version of WebAIR.

The abridged set of WCAG 2.0 guidelines comprised 12 success criteria drawn from each of the three priority levels of WCAG 2.0 (A, AA, and AAA). The WCAG 2.0 index page was modified to display only these success criteria. The associated How to Meet and Understanding pages for each success criteria remained the same.

The abridged version of WebAIR comprised 86 questions across the 10 content categories (a 58% reduction complete version). This was devised by mapping the 12 success criteria identified in the abridged set of WCAG 2.0 guidelines to equivalent questions in WebAIR.

To control for the appearance of the resources, WebAIR was presented using a similar colour scheme and style to the template used by the W3C to present all standards and specifications, including WCAG 2.0.

Participants completed a short questionnaire eliciting their impressions of the accessibility information resources. This comprised 9 five-point Likert items to rate the resource on a range of characteristics: *usefulness* (where 1 = Very low, 5 = Very high), *ease of use* (1 = Very difficult, 5 = Very easy), *navigability* (1 = Very low, 5 = Very high), *understandability* (1 = Very low, 5 = Very high), *completeness* (1 = Very low, 5 = Very high), *amount of information* (1 = Far too little, 5 = Far too much), *number of items to test* (1 = Far too few, 5 = Far too many), *organisation* (1 = Very unclear, 5 = Very clear), and *likelihood of using* (1 = Very unlikely, 5 = Very likely).

---

[59] http://www.w3.org/TR/WCAG/

[60] The initial prototype of WebAIR used in this study is available at: http://www.cs.york.ac.uk/hci/webair/prototype/

[61] The revised version of WebAIR, used in subsequent studies, is available at: http://www.cs.york.ac.uk/hci/webair/

6.3.4. Procedure

As in the previous study, face-to-face sessions with the participants were arranged at a location convenient to them. In many cases, this was at the participant's workplace, but several interviews were conducted in cafés, restaurants, and at participants' homes. Each session was audio recorded using a mobile device.

Following a brief introduction to the investigation and its aims, participants were familiarised with the method. They were assured that the information they would be providing would be confidential and anonymous and their consent was gained to take audio recordings for later transcription. Participants then read and signed an informed consent form (see Appendix A).

Participants were introduced to the four different types of accessibility information resource. For each resource, the purpose and intention behind each one was explained, and participants were given a brief demonstration of how to use it. Participants were then given the opportunity to use the resources themselves and encouraged to assess the accessibility of a website they had developed. For participants who were unwilling or unable to use their own websites, several demonstration websites were provided with known accessibility problems[62].

Participants were given 10 minutes to work with each accessibility information resource. Once they had finished working with each resource, they completed the questionnaire. Participants then answered several open-ended questions on a range of different topics to further explore their opinions about each resource.

---

[62] The demonstration websites included:

- "WCAG 2.0 Before and After Demonstration": a multi-page resource that shows an inaccessible website and a retrofitted version of this same website (http://www.w3.org/WAI/demos/bad/)
- "Accessible University": a fictional academic website designed to demonstrate a variety of common accessibility problems (https://www.washington.edu/accesscomputing/AU/before.html)
- "Untitled on Purpose-An Inaccessible Page": a page that includes a number of accessibility errors and can be used to test reporting tools (http://accessibility.psu.edu/testbadpage)

Once participants had encountered each accessibility information resource, they were asked some further open-ended questions to establish their overall impressions and to investigate which of the accessibility information resources they were most likely to use in future. The interview schedule can be found in Appendix D.

Once the evaluation was complete and the participant had been debriefed, the participants were thanked for their time and asked whether they would be interested in taking part in further research. The study, which immediately followed the interviews conducted as part of the study presented in Chapter 5, lasted approximately 90 minutes.

6.3.5. Data preparation

The audio recordings of the professional web developers were fully transcribed and analysed in the qualitative data analysis software, QSR NVivo[63]. Sample transcripts from three participants can be found in Appendix G. The data was analysed using a conventional content analysis approach (Hsieh & Shannon, 2005) similar to the previous study (see Section 5.3.5).

Each participant's responses were categorised according to the type of issue on which they had commented (e.g. use of accessibility-specific jargon; content too distributed; provides false sense of security). From this open coding approach, an initial set of 44 codes was developed. Secondly, related codes were grouped into 18 subcategories (e.g. language of resource is easy to read; organisation of resource is logical) Finally, related subcategories were combined to form three broad categories. These related to the three key themes that emerged from the contextual inquiry study, presented in Chapter 4, and on which I based the design of WebAIR: language and terminology of the resources; organisation and structure of the resources; and volume and comprehensiveness of the resources (a list of codes and categories can be found in Appendix G).

As in the mental models study, presented in Chapter 5, the analysis was validated by an independent researcher. As in that study, inter-coder reliability was not undertaken but the researcher coded three of the participants' interview transcripts (approximately 10% of the data) using the same codes developed during my analysis. The unit of analysis in this case was each participant's entire response to each question, to which multiple codes could be applied. Due to the semi-structured nature of the interview, not all

[63] http://www.qsrinternational.com/nvivo/nvivo-products

interviewees answered all questions. A degree of consistency between the researcher's and my application of the codes would then serve to corroborate, if not confirm, the results. Any disagreements between the two coding attempts (of which there were very few) were resolved through discussion.

6.4. Results

Figure 6.3 shows the mean ratings of various characteristics of the four accessibility information resources. Descriptive statistics of the mean ratings are presented in Table 6.2. Note: ratings data were not collected for one participant due to error.

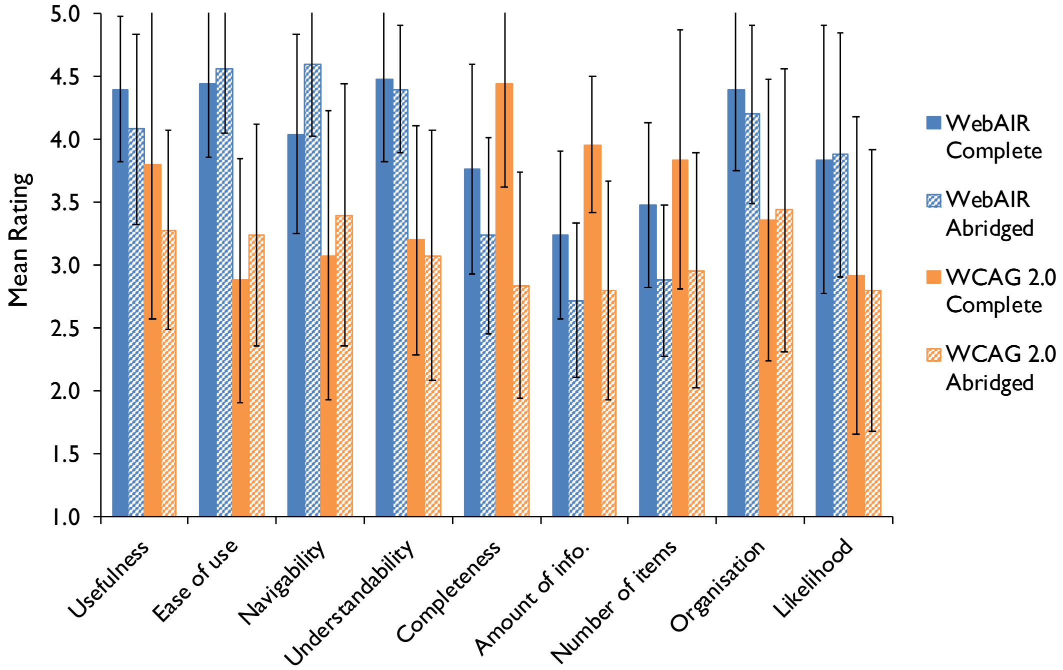


Figure 6.3: Mean ratings of the four accessibility information resources on a range of characteristics

Error bars: standard deviation

A series of two-way repeated measures ANOVAs investigated the effect of *type of resource* (WebAIR or WCAG 2.0) and *size of resource* (complete or abridged versions) on the ratings of the characteristics. While conducting multiple tests may inflate the Type 1 error rate, it nevertheless provides useful comparisons between the different resources.

The tests showed significant effects of *type of resource* on all of the ratings except *completeness*, $F(1, 24) = 1.15$, $p = .295$, $\eta^2_p = .05$, and *number of items to test*, $F(1, 24) = 1.35$, $p = .257$, $\eta^2_p = .05$. WebAIR was rated significantly higher on *usefulness*, $F(1, 24) = 19.60$,

$p < .001$, $\eta^2_p = .45$, *ease of use*, $F(1, 24) = 76.99$, $p < .001$, $\eta^2_p = .76$, *navigability*, $F(1, 24) = 44.18$, $p < .001$, $\eta^2_p = .65$, *understandability*, $F(1, 24) = 53.37$, $p < .001$, $\eta^2_p = .69$, *organisation*, $F(1, 24) = 19.84$, $p < .001$, $\eta^2_p = .45$, and *likelihood of using*, $F(1, 24) = 20.00$, $p < .001$, $\eta^2_p = .46$. WCAG 2.0 was rated significantly higher on *amount of information*, $F(1, 24) = 6.00$, $p = .022$, $\eta^2_p = .20$.

Table 6.2: Descriptive statistics of mean ratings of the four resources on a range of characteristics

Characteristic		***WebAIR Complete***	***WebAIR Abridged***	***WCAG 2.0 Complete***	***WCAG 2.0 Abridged***
Usefulness	***Mean***	4.4	4.1	3.8	3.3
	SD	0.58	0.76	1.22	0.79
Ease of use	***Mean***	4.4	4.6	2.9	3.2
	SD	0.58	0.51	0.97	0.88
Navigability	***Mean***	4.0	4.6	3.1	3.4
	SD	0.79	0.58	1.15	1.04
Understandability	***Mean***	4.5	4.4	3.2	3.1
	SD	0.65	0.50	0.91	1.00
Completeness	***Mean***	3.8	3.2	4.4	2.8
	SD	0.83	0.78	0.82	0.90
Amount of information	***Mean***	3.2	2.7	4.0	2.8
	SD	0.66	0.61	0.54	0.87
Number of items to test	***Mean***	3.5	2.9	3.8	3.0
	SD	0.65	0.60	1.03	0.93
Organisation	***Mean***	4.4	4.2	3.4	3.4
	SD	0.65	0.71	1.11	1.12
Likelihood of using	***Mean***	3.8	3.9	2.9	2.8
	SD	1.07	0.97	1.26	1.12

The tests also showed significant effects of *size of resource* on all of the ratings except *ease of use*, $F(1, 24) = 2.75$, $p = .110$, $\eta^2_p = .10$, *understandability*, $F(1, 24) = 0.86$, $p = .364$, $\eta^2_p = .03$, *organisation*, $F(1, 24) = 0.22$, $p = .641$, $\eta^2_p = .01$, and *likelihood of using*, $F(1, 24) = 0.05$, $p = .828$, $\eta^2_p = .00$. The complete versions of the resources were rated significantly higher on *usefulness*, $F(1, 24) = 6.68$, $p = .016$, $\eta^2_p = .22$, *completeness*, $F(1, 24) = 43.05$, $p < .001$, $\eta^2_p = .64$, *amount of information*, $F(1, 24) = 40.87$, $p < .001$, $\eta^2_p = .63$, and *number of items to test*, $F(1, 24) = 27.24$, $p < .001$, $\eta^2_p = .53$. The abridged versions of the resources were rated significantly higher on *navigability*, $F(1, 24) = 5.76$, $p = .024$, $\eta^2_p = .19$.

There was also a significant interaction between *type* and *size of resource* on ratings of *completeness*, $F(1, 24) = 12.10$, $p = .002$, $\eta^2_p = .34$, and *amount of information*, $F(1, 24) = 9.54$, $p = .005$, $\eta^2_p = .28$. In both cases, this indicates that the effect of *size of resource* (where the complete versions of the resources were rated significantly higher than the abridged versions) was more pronounced for WCAG 2.0 than for WebAIR. This may be due to a more obvious perceived reduction in size in the abridged version of WCAG 2.0 than in that of WebAIR. In other words, both the complete and abridged versions of WebAIR were perceived to be similar in terms of completeness and amount of information. There were no significant interactions between *type* and *size of resource* for any of the other ratings.

Attitudes towards the different resources were elicited by a series of open-ended questions on a range of different topics. Responses to these questions are discussed in the following sections.

6.4.1. Language and terminology of the accessibility information resources

Participants rated both the understandability and ease of use of WebAIR more highly than WCAG 2.0. They were asked to comment on the understandability of the language used in the accessibility information resources.

Nineteen of the twenty-six participants (73%) commented that they struggled with the language of WCAG 2.0. Five participants (19%) acknowledged that technical standards documents such as WCAG 2.0 need a certain degree of precision to be workable and enforceable. Ten participants (38%), however, felt such specificity comes at the expense of readability. Adjectives commonly used by participants to describe the language of WCAG 2.0 were: “wordy”, “verbose”, “academic” and “legalistic”. One participant summarised WCAG 2.0 as having “fairly large run-on sentences for the sake of being very precise but it has the expense of fairly verbose sentences which are a bit difficult to follow” [WD1].

As well as being precise, WCAG 2.0 also presents a lot of information that is specific to the domain of web accessibility and which is often unfamiliar to web developers. Six participants (23%) complained that WCAG 2.0 does not clearly explain domain-specific principles, such as presenting content in a ‘meaningful sequence’ or providing ‘a contrast ratio of at least 4.5:1' between text and its background. One participant (4%) summarised this as: "It feels like you need a good understanding of accessibility before you can approach this, otherwise you will get lost" [WD5]. Although WCAG 2.0 provides a hyperlinked glossary of terms in an appendix, nine participants (35%) complained that linking to and from the glossary disrupted their reading comprehension.

Seventeen participants (65%) stated that they were comfortable with the language of WebAIR and found it easier to read than WCAG 2.0. Six participants (23%) felt the emphasis on web development-specific language and terminology improved the readability of WebAIR and would make it easier for web developers to understand and apply to their work. Summarising the language of WebAIR, one participant (4%) said: "It makes sense to us as an audience. This is the terminology that we use all the time. This is our jargon" [WD4].

Seven participants (27%) felt the question-based phrasing of WebAIR was more intuitive to web developers and easier to understand. One participant (4%) said: "A question forces a simple statement: do you do this or not, explain the reason and how you fix it. Of course, there is the risk of subjective interpretation, but there is an example" [WD4]. The questions appeared to prompt a degree of self-reflection in one participant (4%), who felt that the questions forced him to ask himself whether he understood what he was reading. Question-based phrasing was particularly familiar to one participant (4%), who said it is,

> exactly how I work for my job actually. When I'm designing a new feature, there'll be a set of functional acceptance criteria and they're always worded in that same question manner … so these are really familiar to me in that regard [WD1].

Despite orienting the language towards web development-specific terminology and grounding the guidance in concrete, objective questions, six participants (23%) felt that WebAIR was not pitched at an appropriate level for professional web developers and that it still had to find its audience. Three participants (12%) found the definitions, provided in WebAIR's table of contents, of fundamental web development terms, such as forms or links, unnecessary and a little patronising to professional web developers. Reflecting on the language of WebAIR, one participant (4%) said: "Some of the sections were more simplistic than the level I would say is about right. It is targeted at someone who knows slightly less than me, someone slightly less technical than me" [WD7].

6.4.2. Organisation and structure of the accessibility information resources

Participants rated both the clarity of organisation and navigability of WebAIR more highly than WCAG 2.0. They were asked to comment on the organisation and navigability of the accessibility information resources.

Fifteen participants of the twenty-six participants (58%) felt the organisation of WCAG 2.0 made it difficult to extract useful information. Ten participants (38%) described WCAG 2.0 as having a "logical" or "sensible" organisation that may be appropriate for a technical standards document but not for a day-to-day resource. Six participants (23%) felt the organisation of content according to four 'principles of accessibility' (Perceivable, Operable, Understandable and Robust) was meaningless to web developers. One participant (4%) highlighted how: "You could have called the principles 'Section 1-4' and it would have had the same effect. It's a bit abstract" [WD7].

Six participants (23%) criticised the "top-down" organisation of WCAG 2.0 for prioritising theoretical knowledge over practical guidance. This, they felt, results in techniques that provide useful guidance on how to meet the success criteria being buried deep in the document hierarchy. One participant (4%) described the difficulty he encountered with WCAG 2.0:

> I think in code and [with WCAG 2.0] I have to work to find code samples … If I was having to train somebody and it was the first time they'd seen it, they'd have to really dig through to find examples of best practice [WD14].

Five participants (19%) pointed out that information relating to a particular topic is often distributed across various guidelines and success criteria, forcing them to draw information from several locations. One participant (4%) said:

> I can imagine I'd have to read various bits from each different section in order to decide on a small aspect of the website, even something that's just five lines of code, I can imagine have to read a lot of the document to do it [WD1].

Twenty-one of the twenty-six participants (81%) stated that they were satisfied with the organisation of WebAIR and found it more relevant to web developers than WCAG 2.0. The organisation of information according to content type in WebAIR avoids distributing related information throughout the resource. As well as making it easier for web developers to orient themselves within the resource, seventeen participants (65%)

felt this organisation helped them to identify relevant (and irrelevant) sections of the resource. According to one participant (4%), WebAIR "seems more like it was organised by a developer rather than someone who's just very interested in accessibility" [WD17].

Thirteen participants (50%) particularly appreciated the "bottom-up" prioritisation of practical information in WebAIR, which prevents web developers from having to drill down through a complex hierarchical structure to find the information they need. Participants felt this would be particularly useful when a deadline is imminent, and they might not have time to engage in background reading. One participant (4%) explained how:

> I wouldn't know if I was in the 'Robust' section or one of the others [in WCAG 2.0], whereas WebAIR makes more sense because you can think 'right, I'm working on a form, let's quickly go through the form stuff' [WD4].

Although the majority of participants found the organisation of information according to content type in WebAIR useful, not all categories proved meaningful to them. Ten participants (38%) struggled with the 'time-limited content', 'between pages' and 'within pages' categories as they found these terms unusual. Six participants (23%) suggested that a category relating to website navigation would be useful as it was currently unclear whether this was covered by the 'links' category. Some of the categories were difficult for participants to distinguish. For example, six participants (23%) struggled to understand the difference between 'interactive content; and 'multimedia', with one commenting that the latter was: "a dated concept, reminiscent of CD-ROMs from the late 90s" [WD11].

6.4.3. Volume and comprehensiveness of the accessibility information resources

Participants rated the completeness of WCAG 2.0 more highly than WebAIR but rated the amount of information in WCAG 2.0 as being too much. Participants rated both WCAG 2.0 and WebAIR as including too many items to check. They were asked to comment on the amount of information and the number of items to check in the accessibility information resources.

None of the participants questioned the completeness and thoroughness of WCAG 2.0, however, twenty-two of the twenty-six participants (85%) described it as presenting an "exhaustive", "overwhelming" or "overloaded" amount of information. According to twelve participants (46%), the consequence of presenting so much information at once is that it may discourage web developers from incorporate web accessibility into their workflows. One participant (4%) said the amount of information in WCAG 2.0 can "make you feel this is a bit of an insurmountable thing to actually meet these guidelines" [WD10]. Four participants (15%) warned the amount of information in WCAG 2.0 would be particularly off-putting to anyone new to web accessibility.

Six participants (23%) felt the volume of information presented in WCAG 2.0 made it difficult to prioritise the guidance. One participant (4%) said: "There's obviously a lot to take in and the key thing is you can't pick out where the important ones are" [WD10]. Despite WCAG 2.0 being prioritised according to three levels of conformance: A (lowest), AA, and AAA (highest), six participants (23%) were unclear about the meaning of each conformance level and struggled to identify which success criteria related to which conformance level.

Seventeen participants (65%) felt the considerable number of items to test in WCAG 2.0 hampered its effectiveness as a tool for creating and evaluating accessible websites. Nine participants (35%) felt the size of WCAG 2.0 capably reflects the considerable amount of work required to make websites accessible. Seventeen participants (65%), however, felt that using the guidelines would need an unrealistic amount of time and resources, particularly when clients have not made web accessibility a requirement. One participant (4%) said: "If you insist that some of that stuff is done as a requirement, your clients are going to go 'don't bother doing it, we'll take the hit on people with accessibility problems'" [WD8].

Nine participants (35%) felt the number of items to test in WCAG 2.0 would make it difficult to quickly and confidently establish a benchmark with regards to the accessibility of a website. They anticipated having to spend "hours" or even "days" working through WCAG 2.0 to be able to meet a particular level of conformance. One participant (4%) explained:

> Because there's so many techniques I'd have to spend a long time analysing a site before I could confidently say what level it's at … if I had days spare to do a full

> assessment, I could read all of WCAG 2.0 and eventually give a fairly accurate assessment of the website, but I think for realistic timescales that most people are working to, I wouldn't want to spend an hour or two and then try and suggest that the website is accessible because it is a huge resource [WD1].

Seventeen participants of the twenty-six participants (65%) felt WebAIR contained a more appropriate amount of information than WCAG 2.0. One participant (4%) described WebAIR as providing *"enough information to get the gist"* but *"not too much information to make you run away"* [WD2]. Four participants (15%) considered the latter point to be particularly important for anyone new to web accessibility. By presenting only essential information about each accessibility issue, WebAIR attempts to addresses the problem of information overload. Twelve participants (46%) commented on how this improves the readability and comprehension of WebAIR as well as making it easier to prioritise the guidance. The inclusion of code examples in WebAIR proved popular with sixteen participants (62%), though many felt the examples could be better formatted for readability (e.g. by using syntax highlighting). One participant (4%) described WebAIR as being,

> much simpler than the WCAG guidelines and actually tells you how to fix [accessibility problems]. I could see that being easier to follow than the 'how to meet it' [in WCAG 2.0]. It feels more like a developer way of doing things [WD22].

Seventeen participants (65%) were satisfied with the number of items to test in WebAIR, which some described as being a more "manageable" or a "do-able" amount than in WCAG 2.0. Twenty participants (77%) felt both the number of content categories and the number of questions was appropriate and reflected the website components that web developers typically use, as well as the accessibility issues they might encounter. One participant (4%) said: "Each one of those [categories] you could maybe address in one go, just have a quick assessment of a particular site or page, pick one block at a time and do it in a much more manageable chunk" [WD10].

Many of the questions in WebAIR are follow-up, subsidiary questions that only need to be addressed in certain circumstances. Three participants (12%) commented that the question hierarchy made it easier for them to filter relevant sections of WebAIR, which also improved the memorability of the resource. One participant (4%) said: "I can see that checklist isn't huge, so I know I can retain the points in my brain" [WD3].

Despite WebAIR being much smaller than WCAG 2.0, four participants (15%) felt the number of items to test was still too many. These participants were uneasy about the number of times they would have to work through the information in WebAIR, particularly for frequently updated websites. One participant (4%) said: "I think to run through it for every single site, 100+ questions are perhaps too many. You're not going to want to do 100+ questions if you've not spent that long working on it" [WD2]. Many participants' comments indicated that they rarely consider web accessibility until the very end of the development process, when tackling even a relatively smaller information resource such as WebAIR may still seem a daunting task. One participant (4%) said: "I guess that's the nature of the problem: there's a lot of stuff to think about. But yeah, there's quite a lot to go through" [WD7]. These results reinforce the fact that accessibility is a time-consuming activity that cannot easily be reduced.

6.4.4. Abridged versions of the accessibility information resources

The concept of abridged versions of the accessibility information resources received a hesitant response from participants. This concern was reflected in the participants' ratings of the abridged versions of each resource. The completeness and amount of information of both the abridged versions were rated much lower than both the complete versions. Nine participants (35%) commented that the reduction in size made the resources more manageable and less intimidating. Four participants (15%) felt the abridged versions would be particular useful for introducing colleagues and clients to web accessibility. The prospect of completing web accessibility testing in a more reasonable amount of time was enticing to eight participants (31%), one of whom said: "In real life you are very limited on what time you can spend on this, if any, so it's got to be done in the most pragmatic fashion which helps the most users really" [WD1].

In spite of the potential benefits of the abridged versions, thirteen participants (50%) questioned the usefulness of presenting a limited amount of information. These participants were uneasy about what information was missing from the abridged versions. A further four participants (15%) were concerned that by addressing only a subset of accessibility issues, the abridged versions of the accessibility information resources could lull web developers into a false sense of security. One participant said:

> By believing you have solved the main problems that people are going to face, [you might] get lulled into a false sense of feeling that your website is accessible, whereas the secondary or less common problems could be just as savage, just as severe impediments [WD25].

This suggests that efforts to tackle the problem of information overload that web developers currently encounter should focus not on reducing the scope of information, but instead on presenting the information more succinctly and selectively.

6.4.5. Overall impressions of the accessibility information resources

Participants were asked to provide their overall impressions of the different resources and to reflect on the likelihood of using the resources in future.

Despite the various criticisms made of WCAG 2.0, eleven participants (42%) recognised its significant role as an international standard for web accessibility and inclusive design. Though participants generally found the language, organisation, amount of information and number of items to test made WCAG 2.0 difficult to use, none of them doubted the comprehensiveness and authoritativeness of the resource. Participants considered the latter to be particularly important when clients or organisation required them to meet a particular WCAG 2.0 conformance level. One participant (4%) summarised WCAG 2.0 as being "completely and utterly heavyweight" but added "it is a definitive reference of everything you should really be doing" [WD16].

For day-to-day web development, however, when web developers may not have the time, budget or resources to consult such a comprehensive resource, twenty-two participants (85%) considered WebAIR to be a more realistic and pragmatic alternative. Participants generally felt the language, organisation, amount of information and number of items to test made WebAIR easier to use and more relevant to web developers. The difference between WCAG 2.0 and WebAIR was summarised by one participant (4%) who felt the resources served different purposes. He said:

> WCAG 2.0 is for understanding a bit more about the meaning behind it, the principles … but if I wanted something that I could actually use and apply, I'd go to WebAIR because it links together with actual elements that you'd actually have in a website, rather than it just being general principles [WD11].

6.5. Further evaluation of WebAIR with novice web developers

6.5.1. Method

6.5.1.1. Design

A between-participants laboratory study was undertaken with computer science students, who, as previously mentioned, may be classed as novice web developers.

The novice web developers used the version of WebAIR that had been integrated into the i2Web EASI Tool. Their reactions were elicited using a combination of rating items and open-ended questions administered via an online survey platform.

6.5.1.2. Participants

Seven novice web developers, all of them students, took part in the study. Six participants were female, one was male. The age of participants ranged from 23 to 25 years (mean: 24.0). Five participants were from China, one was from Thailand, and one was from India. Participants were a convenience sample from the students at the University of York. Six of the participants were students from the Department of Computer Science taking the MSc. in Human Centred Interactive Technologies. The seventh participant was a student in the Department of Psychology.

The student participants had between 0.5 and 4 years of experience of web development (mean: 2.0 years), though none had been employed in a web development role. Participants reported having previously developed between 2 to 8 websites (mean: 4.0), ranging from 3 to 15 pages in size (mean: 9.0 pages). Each participant reported being fairly familiar with web accessibility with a mean rating of 4.0 on a 5-point scale (SD: 0.00). Participants reported being moderately familiar with WCAG, with a mean rating of 3.4 on a 5-point scale (SD: 0.54).

6.5.1.3. Materials

Novice web developers were presented with the version of WebAIR that had been integrated into the i2Web EASI Tool. This could be accessed via an 'Accessibility Report' pane within the tool (see Figure 6.4).

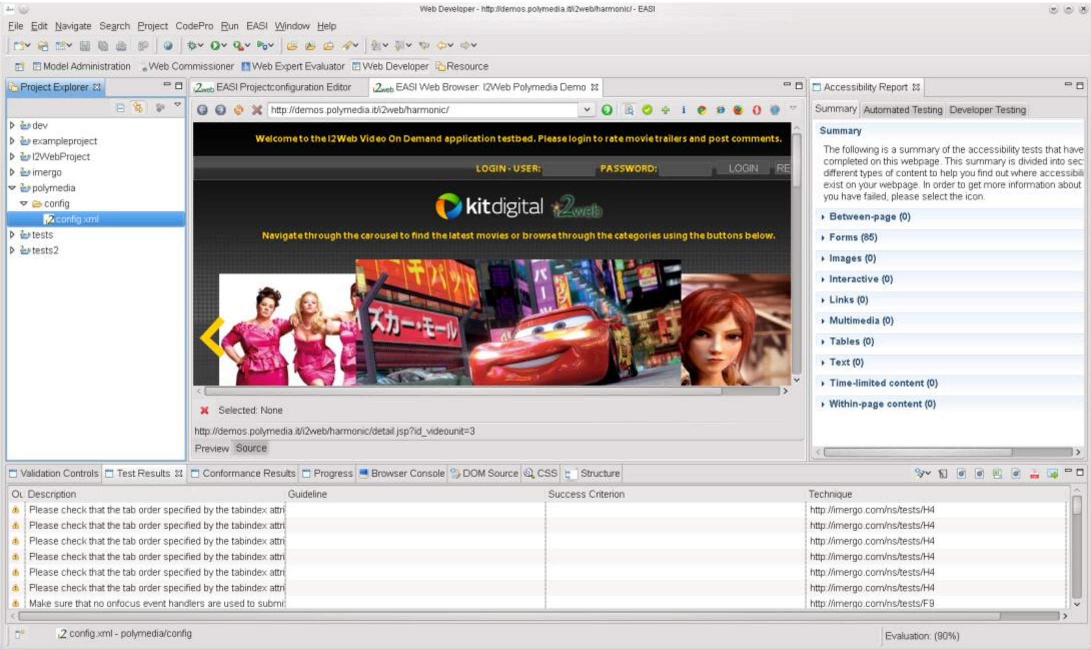


Figure 6.4: EASI Tool main window displaying the 'Accessibility Report' pane. Reprinted with permission.

The ‘Accessibility Report’ pane contains three tabs: a ‘Summary’ tab displaying the number of problems identified by automated and manual testing; an ‘Automated Testing’ tab displaying the pass/fail results of tests that can determined automatically; and a ‘Developer Testing’ tab, displaying the tests that require manual intervention (see Figure 6.5). This report can be generated at any time during or after development.

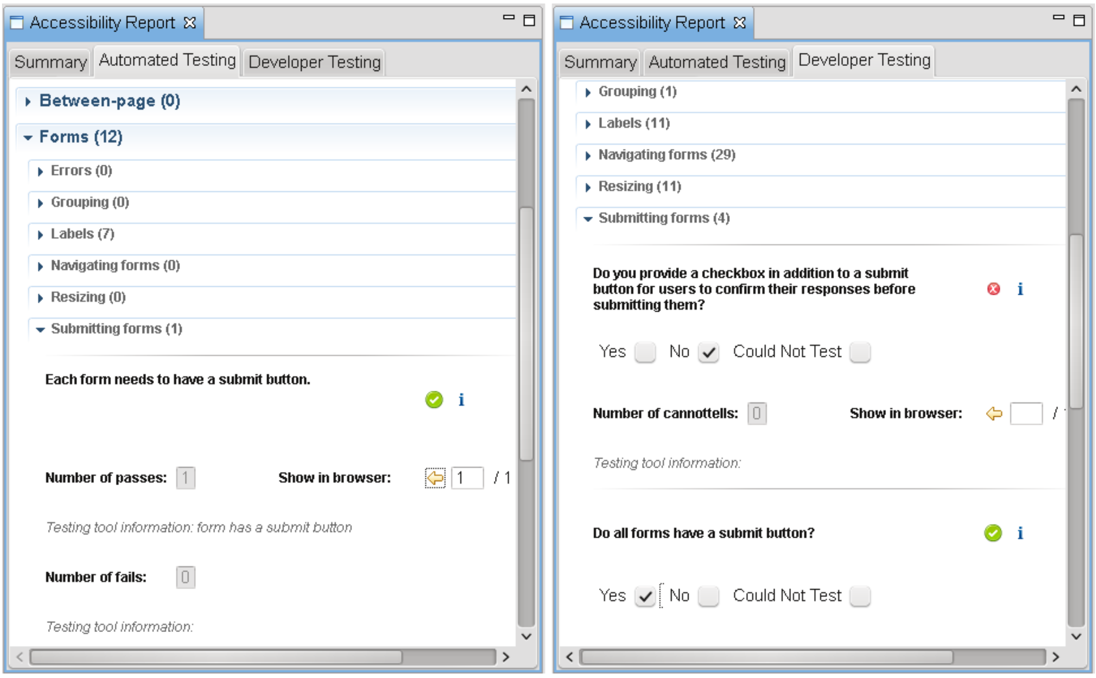


Figure 6.5: ‘Automated Testing’ and ‘Developer Testing’ tabs within the 'Accessibility Report' pane of the EASI Tool. Reprinted with permission.

Each tab adopts the content type categorisation and question-based phrasing of WebAIR. The 'i' (information) icons next to each test link to the 'More Information' pages of WebAIR, which appear within the same pane.

Participants were provided with the EASI Tool version 2.0.1 installed on a Windows 7 personal computer in a laboratory at the University of York. The participants completed a questionnaire that was delivered online through the Google Forms survey platform. They were also given a brief training session on how to use the EASI Tool to evaluate a website. This set of training materials included an example of EASI being used to evaluate a page from Accessibility University, a fictional academic website designed to demonstrate a variety of common accessibility problems[64]. This training document is provided in Appendix G. The participants were also provided with a protocol comprising two evaluation tasks. This protocol is provided in Appendix G. The first task was to conduct an accessibility evaluation of the Accessibility University home page. The second task was to add a second webpage from the Accessibility University website to the sample of pages to be tested and undertake an evaluation across both of the pages. Each task had a small set of questions for the evaluator to answer regarding the accessibility problems on the web pages.

6.5.1.4. Procedure

The study with novice web developers took place in a laboratory at the University of York. Following a brief introduction to the investigation and its aims, participants were familiarised with the method. They were assured that the information they provided would be confidential and anonymous. They then completed an informed consent form (see Appendix A). Following the training session on using the EASI Tool, the participants performed the two evaluation tasks, and answered the questions about the accessibility of the pages they were evaluating. To ensure the evaluation was conducted in a reasonable amount of time, participants were asked to pay particular attention to accessibility errors relating to forms and links only.

Participants then completed an online survey regarding their experiences with the EASI Tool. The survey comprised a set of 5-point Likert items and open-ended questions. As well as various questions about the tool itself, participants were asked about their interaction with WebAIR. The survey is provided in Appendix G.

[64] http://www.washington.edu/accesscomputing/AU/before.html

6.5.2. Results

The following sections summarise the survey findings relevant to the evaluation of WebAIR.

6.5.2.1. Organisation of WebAIR

The mean rating of the clarity of the organisation of WebAIR was 3.7 on a 5-point scale (SD: 0.95, range: 2-5), which was slightly above the midpoint of the scale.

With regards to the WebAIR content categories, one of the seven participants (14%) commented that: "some of the category labels are not very clear: 'between pages', 'time-limited content' and 'within pages'" $_{[NWD7]}$. This echoes the comments made by the professional web developers.

When asked whether there were any content categories that the participants would add or remove, the majority of participants did not provide a response. However, one participant (14%) commented: "For the "interactive" category, it includes some multimedia, which are also included in the "multimedia" categories. I think you should not put the same things in the two different categories" $_{[NWD2]}$.

6.5.2.2. Understandability of WebAIR

The mean rating of the understandability of the WebAIR content was 3.3 on a 5-point scale (SD: 0.49, range 3-4), which was slightly above the midpoint of the scale.

With regards to the understandability of the WebAIR content, one participant (14%) commented: "The examples of the code are not really organised and hard to read" $_{[NWD2]}$. Another participant (14%) said: "the layout of the example is not very clear. All codes in one line" $_{[NWD7]}$. Another participant (14%) felt that some of the documentation was "not sufficient to help" $_{[NWD1]}$. Another (14%) said: "I don't understand the part 'How to Fix the Problem'. In some cases, it is not helpful" $_{[NWD4]}$.

6.5.2.3. Confidence in WebAIR

The mean rating of the participants' confidence in solving accessibility problems using WebAIR was 2.9 on a 5-point scale (SD: 1.07, range 2-5), slightly below the midpoint of the scale.

When asked whether there are any ways in which WebAIR could be improved, participants provided a range of suggestions. Some of these related to the further information links to additional resources.

- "Could provide link to external source" [NWD1].
- "Direct links to further information will be good" [NWD6].
- "If the further help links would be clickable the would perfect" [NWD7].

What is confusing about these responses is that links to further information were provided in the majority of pages in WebAIR. Furthermore, the links had been made clickable. Where further information was not available, this was clearly stated in the help information.

The popularity of the further information may be explained by the following comment from a participant: "[WebAIR] shows how to fix the problem but doesn't show how to do it. So, if user wants to fix a problem, they have to go to further information which I think cannot help that much" [NWD5].

6.5.2.4. Overall impressions of WebAIR

Overall, novice web developers were moderately satisfied with the organisation and understandability of WebAIR and were moderately confidence about their ability to solve accessibility problems using WebAIR. However, several of the participants struggled with how WebAIR was presented. While the code examples in WebAIR were considered useful, participants felt that the way in which the examples had been formatted made them difficult to use. A number of improvements were suggested including adjusting the font size of the help materials, adding more graphic instructions and ensuring that links to additional resources are available from each page of WebAIR.

6.6. Discussion

This chapter presented the design, implementation and initial evaluation of a prototype information resource to support web developers in the creation of accessible websites. The initial prototype of WebAIR was evaluated with 26 professional and 7 computer science students, who were classed as novice web developers. Professional web developers were given the opportunity to try out complete and abridged versions of both WebAIR and WCAG 2.0 in a realistic setting. Novice web developers provided feedback on a version of WebAIR that had been integrated into the EASI Tool, which developed as part of the i2Web project. The objective of both studies was to gain the web developers' initial impressions of the different resources and to measure the impact of different design decisions made in the development of WebAIR on web developers' attitudes and approach to web accessibility.

Both the quantitative and qualitative findings of this study indicate that the design decisions taken in relation to the language and terminology, organisation and structure, and volume and comprehensiveness in WebAIR may be successful in improving the use of web accessibility information for professional web developers, by aligning more closely with their existing working practices.

Professional web developers rated WebAIR more highly across a range of characteristics, considering it to be more useful, easier to use, more navigable, more understandable, and more clearly organised than WCAG 2.0. Though participants considered WCAG 2.0 to be more complete than WebAIR, they felt it provided too much information and too many items to test. Participants felt the amount of information in WebAIR was about right but still found the number of items to test too many. Overall, participants felt more confident about using WebAIR to create accessible websites and, as a result, felt they would be more likely to use WebAIR rather than WCAG 2.0 in future.

Novice web developers, who evaluated a version of WebAIR that had been integrated into the i2Web EASI Tool, rated the resource less positively but provided valuable feedback to guide future development of both WebAIR and the EASI Tool.

The biggest limitation of this study is the duration of time that participants were exposed to each resource. Ten minutes was barely enough time for them to become familiar with each resource, let alone to give them more than a superficial inspection.

Consequently, the study elicited only the participants' first impressions of each resource. Furthermore, the novelty of a new resource (WebAIR), coupled with negative associations with the existing resource (WCAG 2.0), may have unfairly biased participants' responses. Although the participants' time was limited, more considered responses might have emerged from a longer exposure to each resource.

Participants who were aware of my involvement in the design of WebAIR may have felt obliged to provide positive responses. This concern was mitigated to some extent by the assistance of the two other researchers during the data collection. Also, in the evaluations that I conducted, I made efforts to distance myself from WebAIR by describing it as an outcome of the i2Web project to which I was impartial.

Both the previous interview study on mental models and the current evaluation of WebAIR were conducted in the same session over a total of two hours. Although participants were free to take a break at any time, the combination of the two studies may have resulted in fatigue.

6.7. Revision of WebAIR

The initial version of WebAIR incorporated a number of design decisions addressing the requirements that emerged from previous studies. Table 6.3 summarises: the initial/additional requirements; how participants in this initial evaluation received them; and whether they were retained/introduced in the revised version of WebAIR. The subsequent sections explain how the outcome of this initial evaluation informed the development of a revised version of WebAIR[65].

To ensure a consistent editorial style throughout the entire resource, and also to address a number of inaccuracies and typographical errors that had crept into the previous version, I rewrote the content in each of the 'More Information' webpages. This overhaul also gave me the opportunity to introduce content relating to the newly developed HTML 5 standard (Hickson et al., 2014), which, at the time, was still a W3C Candidate Recommendation, and the WAI-ARIA technical specification (Craig & Cooper, 2014), which, at the time, was still a W3C Recommendation.

[65] http://www.cs.york.ac.uk/hci/webair/

Table 6.3: Existing/additional requirements for a web accessibility information resource

ID	*Requirement*	*Reception*	*Retained/Introduced in WebAIR revision?*
RQ1	Must provide specific, unambiguous guidance to web developers.	Participants found the question-based structure of WebAIR more understandable and easier to use. Novice participants found the code examples confusing and difficult to read.	Yes
RQ2	Must use language and terminology that is familiar to web developers.	Participants found the language and terminology of WebAIR easy to read and understand. Some participants considered it too basic for web developers, more appropriate for a general audience.	Yes
RQ3	Must provide a justification or rationale behind the guidance to educate and inform web developers.	Participants did not comment on this aspect of WebAIR.	Yes
RQ4	Must be organised according to web developers' content-oriented workflows.	Participants felt the organisation of WebAIR was logical and easily applicable to their work. Certain categories were considered too abstract or ambiguous to be useful.	Yes
RQ5	Must allow web developers to dip into the guidance as appropriate.	Participants found it easy to navigate by category and quickly gain access to what they needed to do. Participants still see web accessibility as something to be addressed at the end of development.	Yes
RQ6	Should promote user experience over standards conformance.	Participants did not comment on this aspect of WebAIR.	Yes
RQ7	Should encapsulate coverage of WCAG 2.0.	Participants did not comment on this aspect of WebAIR.	Yes
RQ8	Must provide a reasonable and manageable amount of information.	Participants were generally satisfied with the amount of information in WebAIR; enough to get the gist of something but not too much to be off-putting.	Yes
RQ9	Must provide a realistic number of items for web developers to check.	Participants felt WebAIR was manageable but only on an occasional basis. Participants felt WebAIR was too big to be used every time a	Yes

ID	Requirement	Reception	Retained/Introduced in WebAIR revision?
		website is updated. Participants found the abridged version of WebAIR more manageable and easier to use when time is limited. Some participants felt the abridged version of WebAIR could lull web developers into a false sense of security.	
RQ10 (New)	Code examples should be clearly formatted and easy to read.	Both professional and novice participants felt the clarity of the code examples could be improved by using syntax highlighting.	No
RQ11 (New)	Should provide comprehensive solutions to each issue.	Novice participants felt the lack of detail in the 'More Information' pages forced them to unnecessarily consult external resources.	No
RQ12 (New)	Should link to at least one external resource from every 'More Information' page.	Novice participants felt there were too many 'More Information' pages that lacked links to external resources.	Yes

6.7.1. Language and terminology

The phrasing of accessibility issues as specific, unambiguous questions **[RQ1]** proved popular with the majority of participants. Novice web developers did not rate the understandability of WebAIR very highly. However, this was largely due to the presentation of the code examples, which they found confusing and difficult to read. This design decision was retained in the revised version of WebAIR, but a number of questions were removed. This was either due to the question being redundant, ambiguous or, in some cases, incorrect. For example, the first question in the *Forms* category of the initial version of WebAIR asked: *Do all form controls have text labels?* This was followed by the question: *Does each control in the form have a label that describes what it is for?* Both questions essentially asked the same thing and so the latter was removed. Similarly, the question: *Is the text of a sufficient size and contrast?* was asked in both the *Text* and *Links* sections and so, again, the latter was removed. This refactoring process reduced the total number of first-level questions from 112 in the initial version of WebAIR to 86 in the revised version: a reduction of 23% **[RQ9]**. Due to time limitations, the presentation of the code examples was not adjusted **[RQ10]**, however this may be considered in future versions of the resource.

The use of development-centred language and terminology that avoids the domain-specific language of web accessibility **[RQ2]** was also well received by the majority of participants in this study. Some participants considered the language of WebAIR too basic and felt that it laboured the more familiar concepts. In avoiding domain-specific language, there is evidently a danger in over-simplifying the more familiar terminology to the extent that web developers no longer feel it is relevant to them. However, as the resource is intended to appeal to a broad range of web developers, particularly those who are less familiar with web accessibility, this design decision was retained in the revised version of WebAIR.

None of the participants explicitly commented on the inclusion of a rationale to explain why each question is important and what the consequences of not addressing it would be **[RQ3]**. Instead, they tended to focus on whether the resource provided enough information to address each issue. Nevertheless, considering the importance of this requirement in previous studies, this design decision was retained in the revised version of WebAIR.

6.7.2. Organisation and structure

The majority of participants preferred the organisation of information by web content types in WebAIR **[RQ4]**. Categories that did not correspond directly to specific types of content however, such as those relating to time-limited content, between-page and within-page navigation, were considered too abstract or ambiguous to be useful. While many web accessibility issues correspond directly to specific types of content, some are much broader in scope and relate to an entire webpage or website. While this design decision was retained, some of the categories were renamed and others were combined. For example:

- *Multimedia* was renamed *Audio & Video* to better reflect its relation to audio and video content on webpages;
- *Forms* was renamed *Forms & Controls* to incorporate all interactive components on webpages;
- *Time-limited content* was renamed *Time Limits* to clarify its purpose;
- *Within pages* and *Between pages* were combined to form a single *Navigation* category; and

- *Interactive Content* was removed, and its content distributed between the *Audio & Video* and *Forms* categories.

This re-organisation of the content categories reduced the total number from 10 to 8 but, hopefully, results in a more intuitive and concrete categorisation. Further work is necessary to identify appropriate organisational structures for orienting web developers within the information resource for more global webpage issues.

The organisation of WebAIR into discrete categories was designed to allow web developers to dip into the guidance as appropriate **[RQ5].** The aim of this was to encourage the use of WebAIR throughout the web development process. Although participants did not explicitly comment on this, their comments about the amount of work that WebAIR requires, suggests that they still see web accessibility as something to be addressed at the end of development. While this design decision was retained in the revised version of WebAIR, further work is needed to better integrate accessible design and iterative testing into development tasks and tools.

None of the participants explicitly commented on the absence of conformance or priority levels in WebAIR, which was intended to promote user experience over standards conformance **[RQ6]**. This is perhaps to be expected given the limited time available to evaluate each resource. Nevertheless, given the importance of this approach for promoting a more holistic understanding of web accessibility, this design decision was retained in the revised version of WebAIR.

6.7.3. Volume and comprehensiveness

None of the participants explicitly commented on WebAIR's coverage of existing guidelines. However, a number of participants were concerned about the information missing from the abridged version of the resource. This supports the requirement that the resource should encapsulate coverage of WCAG 2.0 **[RQ7]** and suggests that efforts to reduce the scope of information are less effective than presenting the information more succinctly and selectively.

As the abridged versions of each resource were poorly received by participants, I decided not to pursue this alternative approach to reducing the amount of information and the number of items to check **[RQ8, RQ9]** in the revised version of WebAIR.

Despite WebAIR being much smaller than the WCAG 2.0 documentation **[RQ8, RQ9]**, participants still felt it was too large. However, the design decision to include only one example solution drawn from the numerous techniques provided in WCAG 2.0 left very few options for further reducing the number of items to test in the revised version of WebAIR. Also, novice participants felt the lack of detail in the 'More Information' pages forced them to unnecessarily consult external resources **[RQ11]**. Consequently, this design decision was retained and the amount of information in WebAIR remained roughly the same. Links to external resources were included wherever appropriate **[RQ12]**.

6.8. Conclusions

WebAIR is a promising prototype information resource that was designed with web developers in mind. The purpose of WebAIR was to help web developers in learning about web accessibility and support them in the creation and evaluation of accessible websites. It was designed to present web developers with a digestible amount of accessibility support material and accessible web development techniques that are written in web development-oriented language and organised around their existing working practices.

The outcome of this initial evaluation was very encouraging: professional web developers rated WebAIR more highly across a range of characteristics, considering it to be more useful, easier to use, more navigable, more understandable, and more clearly organised than WCAG 2.0. Novice web developers, though less positive about WebAIR, provided valuable feedback to guide future development of the resource.

I used the considerable amount of feedback generated by this evaluation and the many suggestions for improvement made by the web developers to guide a substantial revision and rewrite of WebAIR. This also incorporated the outcome of the mental models interview study that I had conducted in parallel to this initial evaluation.

As well as marking the point at which I was no longer developing the resource under the auspices of the i2Web project, this also signalled the shift into the evaluation stage of this investigation. In the studies presented in the following chapters, I evaluate the effectiveness of the revised version of WebAIR in supporting web developers to create and evaluate accessible websites.

Chapter 7. Evaluation of WebAIR's effectiveness in supporting web accessibility evaluation

7.1. Introduction

The evaluation of WebAIR presented in the previous chapter assessed web developers' initial impressions of the resource. The encouraging results of this study indicated that the design decisions made in the development of the resource had been effective and that WebAIR would be well received by both professional and novice web developers. Despite the previous study being conducted in realistic settings (typically, the participants' workplace), participants' exposure to the resource was limited. Consequently, participants barely got the opportunity to use the resource for its intended purpose: to support them in the creation and evaluation of accessible websites. To further evaluate WebAIR's effectiveness for this purpose, a more rigorous, controlled experimental design was necessary.

This chapter presents a laboratory-based evaluation of WebAIR with computer science students, who are novice web developers. As in the previous chapter, despite one of the aims of WebAIR being to further web developers' knowledge and understanding of web accessibility, this aspect of the resource could not adequately be evaluated in this study. Although the amount of time that participants were exposed to each resource was slightly increased to 25 minutes, it was still not sufficient to observe any substantial changes in knowledge or understanding. Instead, the study objective was to evaluate the ease of use and effectiveness of WebAIR in supporting novice web developers to confidently identify accessibility problems and solutions in existing websites.

7.2. Method

7.2.1. Design

To investigate the effectiveness of WebAIR in supporting novice web developers to evaluate the accessibility of websites, a within-participants study with 48 computer science students at the University of York was conducted. Participants alternatively used both WebAIR and WCAG 2.0 to identify accessibility problems and solutions in the homepages of two publicly available websites.

The choice of websites to evaluate was based on a number of factors. To expose participants to different sections of the resources, the websites included a variety of content types (e.g. text, images, forms, audio and video etc.) and exhibited a variety of accessibility problems. Instead of evaluating fictional, artificial demonstration websites, such as those used in the previous study, participants evaluated modern, professional, real-life websites. Any disadvantages over the potential lack of control and predictability over live websites would be outweighed by greater ecological validity, with regards to the authenticity and plausibility of the task, and increased appeal of the task to the student participants. Following an extensive search of Internet message boards, forums, and mailing lists for websites that would meet these criteria, the websites of Manchester United football club[66] (see Figure 7.1) and the Pure Gym gym operator[67] (see Figure 7.2) were chosen. Both webpages, at the time, had attracted criticism on Internet forums and social media for being inaccessible to people with disabilities.

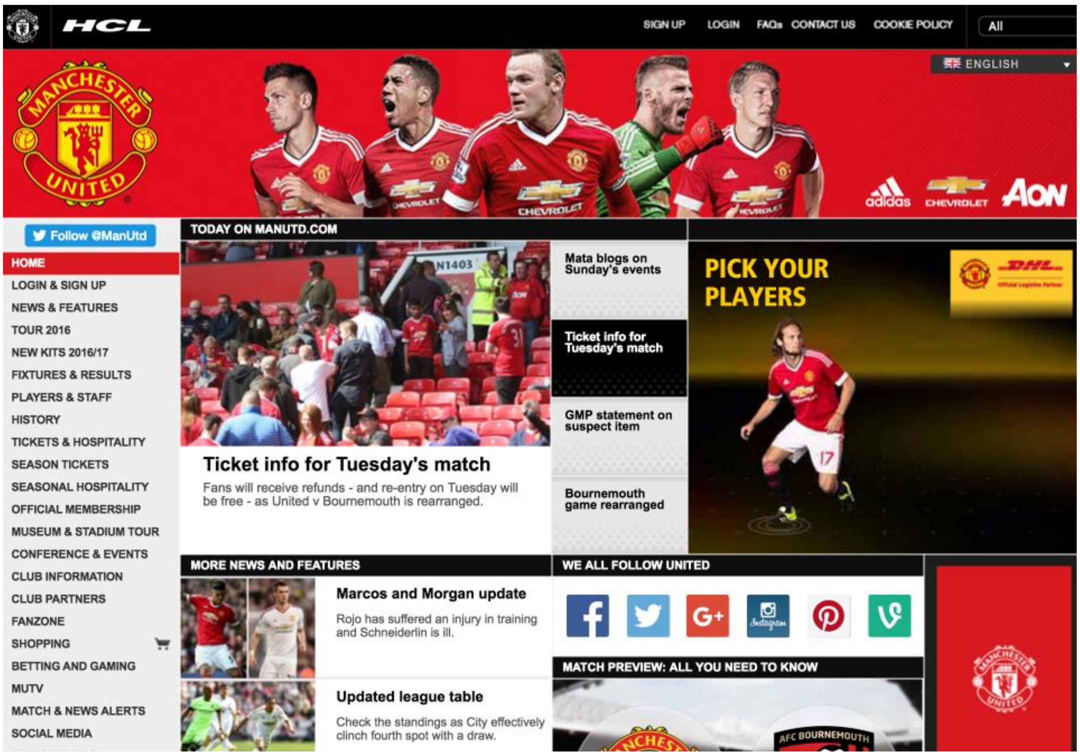


Figure 7.1: Screenshot of Manchester United football club homepage. Reprinted with permission.

[66] http://www.manutd.com
[67] http://www.puregym.com

The student participants were all taking the Interactive Application Programming Techniques (IAPT) module run by the Department of Computer Science. The aim of the IAPT module is to provide an understanding of human computer interaction theories and concepts that form the foundation of advanced interactive systems. At the point in the IAPT module when the study was conducted, the students had only had a brief introduction to web accessibility. Considering both the students' inexperience and the limited amount of time available in which to conduct the study (approximately two hours), the module leader (Dr Christopher Power) and I felt it would be neither appropriate nor useful for the students to conduct full web accessibility evaluations. Instead, the students answered a series of questions that would guide them towards identifying potential accessibility problems in the evaluation webpages. Though the artificial nature of the evaluation may have threatened the ecological validity of the study, it represented a necessary trade-off between evaluating the accessibility information resources effectively and providing a useful learning experience for the students involved.

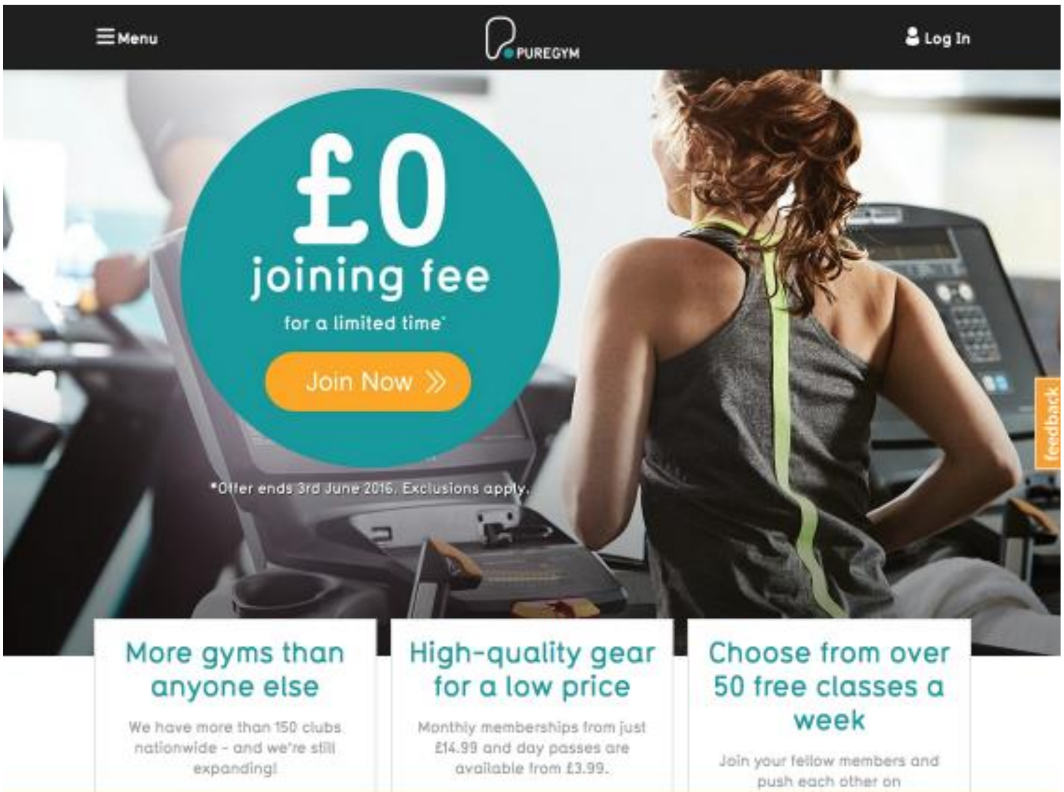


Figure 7.2: Screenshot of Pure Gym gym operator homepage. Reprinted with permission.

7.2.2. Participants

Forty-eight students voluntarily participated in the study. Four participants were female, the rest were male. The age of participants ranged from 20 to 24 years (mean: 22.0). Twenty-eight participants were from the UK; six participants were from Bulgaria; five participants were from Romania. The remaining nine participants were from: Moldova, Jersey, Latvia, Singapore, Estonia, China, Ghana, Venezuela, and Lithuania. All participants were studying for bachelors or masters qualifications in computer science or a related discipline at the University of York. Thirty-four participants were studying for their first degree. The remainder had a previous degree in computer science. Participants did not receive any compensation for their involvement in the study. However, the module leader (Dr Christopher Power) considered the study to be a relevant learning experience from which the students could benefit and be able to use in their module assessment. Informed consent was obtained from all participants.

Participants were assigned to one of four counter-balanced groups, which differed on order of presentation, both of the two evaluation webpages (Manchester United and Pure Gym homepages) and of the two accessibility information resources (WebAIR and WCAG 2.0). Of the 58 participants who were initially present, nine provided incomplete data and were later excluded from the study. One participant did not consent to take part. This resulted in slightly unequal group sizes of 12, 10, 14 and 12 participants.

Participants' familiarity with developing websites varied. Overall, participants reported having between 0 and 8 years' experience of web development (mean: 1.9 years; SD: 2.13). Seven participants had previously been employed as web developers. Participants reported having previously developed between 0 and 20 websites (mean: 2.8 websites; SD: 3.39). Participants estimated that the websites they had previously developed had between 1 and 60 pages (mean: 10.4 pages; SD: 10.05).

On scales of 1 to 5 (1 = very unfamiliar; 5 = very familiar), participants reported being fairly unfamiliar with web accessibility (mean: 2.6; SD: 0.89) and very unfamiliar with WCAG 2.0 (mean: 1.5; SD: 0.58). These ratings are understandable, given that the participants had only recently been introduced to web accessibility as part of the IAPT module.

7.2.3. Materials

The study was conducted in a computer laboratory in the Computer Science department at the University of York. All participants had access to a Windows 7 personal computer with Internet access. Participants' access to their own Internet-enabled smartphones, tablets or other devices was unrestricted and some chose to use these devices as second screens on which to view the study materials.

To guide participants towards identifying potential accessibility problems, they were presented with a set of questions for each of the two evaluation webpages (Manchester United and Pure Gym homepages). Each question asked participants whether or not a particular component of the webpage could cause problems for people with disabilities. For example, one of the questions about the Manchester United homepage asked: *Could the images in the 'More News & Features' section cause any problems for people with disabilities?* The questions required participants to describe both the problem and its solution and to rate how confident they were in their response on a scale of 1 to 5 (1 = not at all confident; 5 = extremely confident).

To ensure that participants would consider each question carefully, a number of 'red herring' questions were included for which an accessibility problem did not exist in the evaluation webpages. For example, one of the questions about the Pure Gym homepage asked: *Could people with disabilities experience any problems if they do not provide all of their details in the 'Fill In Your Details' form?* The form in question is validated upon submission and, in accordance with WCAG 2.0 Success criterion 3.3.2 (Labels or Instructions), any empty required fields are properly indicated. Consequently, this particular component was not considered problematic for people with disabilities.

Fully navigating and exploring the resources would be unfeasible in the limited time available, so, to guide participants' navigation and to focus their identification of accessibility problems and solutions, they were presented with a list of common accessibility recommendations and where to find the relevant guidance in both WebAIR and WCAG 2.0. The recommendations included: the use of appropriate, semantic HTML markup; the use of text alternatives for all non-text content; the clear labelling of all links, buttons, form controls and other interactive components; keyboard navigability of all webpage elements; and the provision of clear feedback.

To elicit participants' impressions of the accessibility information resources, the same questionnaire used in the previous study was administered (see Section 6.3.3).

Participants were presented with a pack of paper materials containing: a consent form; a demographic questionnaire; the two sets of questions relating to potential accessibility problems in the two evaluation webpages; the list of references to common accessibility recommendations in the two resources; and the two questionnaires for evaluating the resources. They were also given access to a pre-formatted spreadsheet via Google Drive to allow them to enter their responses to the study questions. A copy of each of these materials can be found in Appendix G.

7.2.4. Procedure

The study was conducted during a two-hour practical session in the IAPT module. To introduce participants to the purpose of the study, I delivered a short PowerPoint presentation including a demonstration of how to evaluate different aspects of webpages using various web browsers (Mozilla Firefox, Google Chrome, Microsoft Internet Explorer, and Apple Safari). This included an example of how to use each of two accessibility information resources to address an accessibility problem (but not one of the accessibility problems that they might encounter in the evaluation webpages). Following the introductory presentation, participants completed the first part of the consent form.

Participants were given five minutes to familiarise themselves with the accessibility information resource they had been allocated, and to ask any questions about the study. They were assured that, given the number of questions for each webpage and the size of each accessibility information resource, they should not feel they had to attempt all of the questions. It was emphasised that the purpose of the study was to investigate the effectiveness of the accessibility information resources and that participants should not worry if they did not feel very confident about the evaluation.

Participants were then given 25 minutes to attempt as many of the study questions as possible using the accessibility information resource they had been allocated, and to log their responses on the spreadsheet provided. Two of the groups used WebAIR for the first half of the session; the other two groups used WCAG 2.0. Two of the groups evaluated the Manchester United homepage for the first half of the session; the other two groups evaluated the Pure Gym homepage. Participants then completed a questionnaire relating to the accessibility information resource they had just been using.

Participants then swapped both the accessibility information resource and evaluation webpage for the second half of the session. Again, they were given five minutes to familiarise themselves with the accessibility information resource they had been allocated. They were then given another 25 minutes to attempt as many of the study questions as possible using the accessibility information resource they had been allocated, and to log their responses on the spreadsheet provided. After 25 minutes, participants were asked to stop what they were doing and to complete a questionnaire relating to the accessibility information resource they had just been using.

Finally, participants completed the demographics questionnaire and the second part of the online consent form, in which they provided their consent to submit their responses. I concluded the session by delivering a short PowerPoint presentation, which served to debrief the participants and provide them with a demonstration of how they could go on to find accessibility problems as part of a full web accessibility evaluation.

7.2.5. Data preparation

To allow a by-participant analysis of the data, three measures of performance were derived for each participant: productivity (the number of questions attempted, regardless of accuracy); accuracy (the percentage correct of the number of questions attempted); and confidence (the mean self-rating of confidence for all questions attempted). Separate measures of productivity and accuracy were derived for both the problem and solution aspect of each question. This procedure resulted in five dependent variables: *productivity* (in relation to problems); *accuracy* (in relation to problems); *productivity* (in relation to solutions); *accuracy* (in relation to solutions); and *confidence*.

To allow a by-item analysis of the data, two performance measures were derived for each question: productivity (the number of participants who attempted the question, regardless of accuracy); and accuracy (the number of correct responses to the question). Separate measures of productivity and accuracy were derived for both the problem and solution aspect of each question. This resulted in a further four dependent variables:

productivity (in relation to problems); *accuracy* (in relation to problems); *productivity* (in relation to solutions); and *accuracy* (in relation to solutions), To avoid skewing the results, questions that had been attempted by less than a third (16) of all participants were excluded. This left a remainder of seven questions relating to the Manchester United homepage and eight questions relating to the Pure Gym homepage that were suitable for analysis.

7.3. Results

Figure 7.3, below, shows the mean ratings of various characteristics of WebAIR and WCAG 2.0. Descriptive statistics of the mean ratings are presented in Table 7.1.

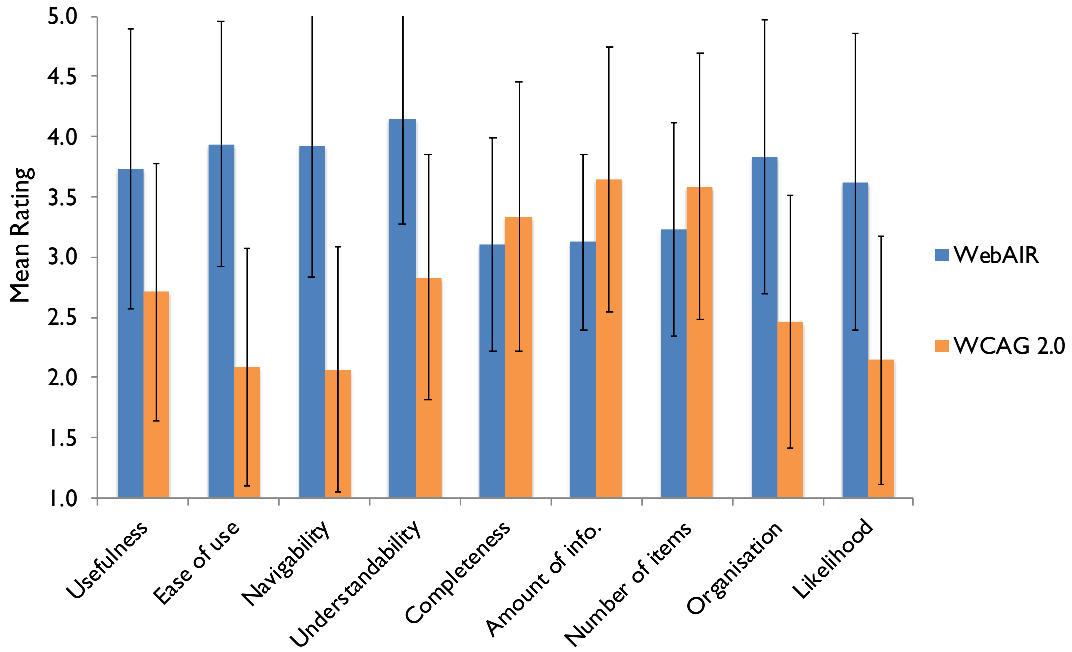


Figure 7.3: Mean ratings of WebAIR and WCAG 2.0 on a range of characteristics

Error bars: standard deviation

A series of one-way repeated measures ANOVAs investigated the effect of *type of resource* (WebAIR or WCAG 2.0) on the ratings of the characteristics. Again, while conducting multiple tests may inflate the Type 1 error rate, it nevertheless provides useful comparisons between the different resources.

The tests showed significant effects of *type of resource* on all of the ratings except *completeness*, $F(1, 47) = 1.18$, $p = .283$, $\eta^2_p = .02$. WebAIR was rated significantly higher on *usefulness*, $F(1, 47) = 25.84$, $p < .001$, $\eta^2_p = .36$, *ease of use*, $F(1, 47) = 107.75$, $p < .001$, $\eta^2_p = .70$, *navigability*, $F(1, 47) = 92.36$, $p < .001$, $\eta^2_p = .66$, *understandability*, $F(1, 47) =$

55.27, $p < .001$, $\eta^2_p = .54$, *organisation*, $F(1, 47) = 46.74$, $p < .001$, $\eta^2_p = .50$, and *likelihood of using*, $F(1, 47) = 44.08$, $p < .001$, $\eta^2_p = .48$. WCAG 2.0 was rated significantly higher on *amount of information*, $F(1, 47) = 8.27$, $p = .006$, $\eta^2_p = .15$, and *number of items to test*, $F(1, 47) = 4.23$, $p = .045$, $\eta^2_p = .08$.

Table 7.1: Descriptive statistics of mean ratings of WebAIR and WCAG 2.0 on a range of characteristics

Characteristic		***WebAIR***	***WCAG 2.0***
Usefulness	***Mean***	3.7	2.7
	SD	1.16	1.07
Ease of use	***Mean***	3.9	2.1
	SD	1.02	0.99
Navigability	***Mean***	3.9	2.1
	SD	1.09	1.02
Understandability	***Mean***	4.1	2.8
	SD	0.87	1.02
Completeness	***Mean***	3.1	3.3
	SD	0.88	1.12
Amount of information	***Mean***	3.1	3.6
	SD	0.73	1.10
Number of items to test	***Mean***	3.2	3.6
	SD	0.88	1.11
Organisation	***Mean***	3.8	2.5
	SD	1.14	1.05
Likelihood of using	***Mean***	3.6	2.1
	SD	1.23	1.03

To investigate the possibility of a training effect, a series of two-way mixed model repeated measures ANOVAs investigated the effect of *exposure* (WebAIR first or WCAG 2.0 first) and *type of resource* (WebAIR or WCAG 2.0) on the ratings of the characteristics. This revealed no significant interaction between *exposure* and *type of resource* for any of the ratings. This suggests that the ratings were not affected by the order in which participants were exposed to the resources.

The following sections summarise the results of the accessibility evaluation tasks that participants performed on the Manchester United and Pure Gym homepages.

7.3.1. Manchester United homepage accessibility evaluation task

Descriptive statistics of the participants' performance in the Manchester United homepage accessibility evaluation task are presented in Table 7.2.

Table 7.2: Descriptive statistics of participants' performance in the Manchester United homepage accessibility evaluation task

		WebAIR	***WCAG 2.0***
	N	26	22
Productivity (Problems)	***Mean***	6.46	6.27
	SD	3.127	2.640
	Min	1	3
	Max	16	14
Accuracy (Problems)	***Mean***	63%	67%
	SD	0.243	0.265
	Min	0%	100%
	Max	0%	100%
Productivity (Solutions)	***Mean***	5.12	4.41
	SD	2.732	2.443
	Min	1	1
	Max	14	10
Accuracy (Solutions)	***Mean***	52%	60%
	SD	0.268	0.343
	Min	0%	100%
	Max	0%	100%
Confidence	***Mean of means***	3.59	3.28
	SD	0.696	0.657
	Min	2.13	1.71
	Max	4.71	4.50

Because the data failed the normality assumption for ANOVA, a series of one-way non-parametric Kruskal-Wallis tests investigated the effect of *type of resource* (WebAIR or WCAG 2.0) on the overall performance measures. Due to the presence of outliers in some of the dependent variables, the analyses were repeated with Winsorised data at the .05 level (i.e. outliers were replaced with the 5th or 95th percentile values, depending on the tail of the distribution).

Tests on both the non-Winsorised and Winsorised data revealed no significant effect of *type of resource* on any of the dependent variables, including *productivity* (in relation to problems) (Non-Winsorised: $\chi 2(1) = 0.63$, $p = .802$; Winsorised: $\chi 2(1) = 0.63$, $p = .802$), *accuracy* (in relation to problems) (Non-Winsorised: $\chi 2(1) = 0.190$, $p = .663$; Winsorised: $\chi 2(1) = 0.021$, $p = .884$), *productivity* (in relation to solutions) (Non-Winsorised: $\chi 2(1) = 1.303$, $p = .254$; Winsorised: $\chi 2(1) = 1.327$, $p = .249$), *accuracy* (in relation to solutions) (Non-Winsorised: $\chi 2(1) = 0.749$, $p = .387$; Winsorised: $\chi 2(1) = 0.749$, $p = .387$), and *confidence* (Non-Winsorised: $\chi 2(1) = 2.616$, $p = .106$; Winsorised: $\chi 2(1) = 2.616$, $p = .802$). These results indicate that participants were no more productive, accurate or confident in their responses using either resource.

A series of Chi Square analyses compared participants' productivity and accuracy in the seven most frequently attempted questions in the Manchester United homepage accessibility evaluation task. The small number of participants attempting some of the questions violated the cell size assumption of the Chi Square test. Where the expected frequencies were less than five, Fisher's exact test was calculated instead. Results of the analyses are presented in Table 7.3 and Table 7.4, respectively. The analyses revealed no significant differences in either the *productivity* or *accuracy* measures between the two conditions in relation to either problems or solutions for any of the questions. These results indicate that participants were no more or no less productive or accurate in any of the seven most frequently attempted questions using either resource.

Descriptive statistics of the participants' self-rating of *confidence* in their responses to the seven most frequently attempted questions in the Manchester United homepage accessibility evaluation task are presented in Table 7.5. A further series of one-way non-parametric Kruskal-Wallis tests investigated the effect of *type of resource* (WebAIR or WCAG 2.0) on the mean self-rating of *confidence* for each individual question. The tests revealed a significant effect of *type of resource* on participants' confidence on two of the questions: Question 1 ($\chi 2(1) = 5.069$, $p = .024$) and Question 4 ($\chi 2(1) = 4.098$, $p = .043$). For both of these questions, participants were significantly more confident in their responses when using WebAIR than when using WCAG 2.0. There was no significant effect of *type of resource* on participants' confidence for any of the other questions. These results indicate that, with the exception of Question 1 and Question 4, participants were no more or no less confident in their responses to any of any of the seven most frequently attempted questions using either resource.

Table 7.3: Frequencies and Chi Square test/Fisher's Exact test results relating to participants' productivity in the seven most frequently attempted questions in the Manchester United homepage accessibility evaluation task

			WebAIR		*WCAG 2.0*		*Total*			
		N	***Attempted***	***Not attempted***	***Attempted***	***Not attempted***	***Attempted***	***Not attempted***	***χ(1)***	***P-Value***
Q1	***Problem***	48	25	1	21	1	46	2	N/A	1.000
	Solution	48	24	2	16	6	40	8	N/A	0.119
Q2	***Problem***	48	24	2	20	2	44	4	N/A	1.000
	Solution	48	18	8	14	8	32	16	0.168	0.682
Q3	***Problem***	48	22	4	20	2	42	6	N/A	0.674
	Solution	48	16	10	11	11	27	21	0.645	0.422
Q4	***Problem***	48	22	4	17	5	39	9	N/A	0.713
	Solution	48	17	9	13	9	30	18	0.201	0.654
Q5	***Problem***	48	19	7	18	4	37	11	0.515	0.473
	Solution	48	15	11	15	7	30	18	0.559	0.454
Q6	***Problem***	48	11	15	14	8	25	23	2.172	0.141
	Solution	48	9	17	9	13	18	30	0.201	0.654
Q7	***Problem***	48	12	14	10	12	22	26	0.002	0.961
	Solution	48	11	15	8	14	19	29	0.176	0.675

Table 7.4: Frequencies and Chi Square test/Fisher's Exact test results relating to participants' accuracy in the seven most frequently attempted questions in the Manchester United homepage accessibility evaluation task

			WebAIR		*WCAG 2.0*		*Total*			
		N	*Correct*	*Incorrect*	*Correct*	*Incorrect*	*Correct*	*Incorrect*	*χ(1)*	*P-Value*
Q1	***Problem***	46	17	8	14	7	31	15	0.009	0.923
	Solution	40	15	9	9	7	24	16	0.156	0.693
Q2	***Problem***	44	12	12	10	10	22	22	0.000	1.000
	Solution	32	12	6	10	4	22	10	N/A	1.000
Q3	***Problem***	42	12	10	10	10	22	20	0.087	0.768
	Solution	27	2	14	0	11	2	25	N/A	0.499
Q4	***Problem***	39	15	7	13	4	28	11	N/A	0.725
	Solution	30	14	3	10	3	24	6	N/A	1.000
Q5	***Problem***	37	17	2	17	1	34	3	N/A	1.000
	Solution	30	14	1	14	1	28	2	N/A	1.000
Q6	***Problem***	25	8	3	10	4	18	7	N/A	1.000
	Solution	18	7	2	6	3	13	5	N/A	1.000
Q7	***Problem***	22	11	1	9	1	20	2	N/A	1.000
	Solution	19	1	10	4	4	5	14	N/A	0.111

Table 7.5: Descriptive statistics relating to participants' confidence in their responses to the seven most frequently attempted questions in the Manchester United homepage accessibility evaluation task

	WebAIR				*WCAG 2.0*					
	N	*Mean*	*Std. Dev.*	*Min.*	*Max.*	*N*	*Mean*	*Std. Dev.*	*Min.*	*Max.*
Q1	25	3.80	0.957	2	5	21	3.10	1.044	1	5
Q2	24	3.29	1.301	1	5	20	3.60	1.046	1	5
Q3	21	3.29	1.189	1	5	19	3.37	1.012	1	5
Q4	22	3.91	1.151	1	5	16	3.19	1.223	1	5
Q5	18	3.94	1.162	1	5	16	3.44	1.263	1	5
Q6	11	3.82	1.079	2	5	12	3.33	1.155	1	5
Q7	12	3.75	1.288	2	5	8	4.13	1.126	2	5

7.3.2. Pure Gym homepage accessibility evaluation task

Descriptive statistics of the participants' performance in the Pure Gym homepage accessibility evaluation task are presented in Table 7.6.

Table 7.6: Descriptive statistics of participants' performance in the Pure Gym homepage accessibility evaluation task

		WebAIR	***WCAG 2.0***
	N	22	26
Productivity (Problems)	***Mean***	7.00	6.35
	SD	3.039	2.884
	Min	3	3
	Max	16	14
Accuracy (Problems)	***Mean***	59%	47%
	SD	0.165	0.216
	Min	0%	0%
	Max	100%	100%
Productivity (Solutions)	***Mean***	5.14	4.46
	SD	2.569	2.404
	Min	3	0
	Max	14	10
Accuracy (Solutions)	***Mean***	48%	32%
	SD	0.265	0.235
	Min	0%	0%
	Max	100%	100%
Confidence	***Mean of means***	3.62	3.53
	SD	0.723	0.614
	Min	2.00	1.88
	Max	4.44	4.25

Again, because the data failed the normality assumption for ANOVA, a series of one-way non-parametric Kruskal-Wallis tests investigated the effect of *type of resource* (WebAIR or WCAG 2.0) on the overall performance measures. Due to the presence of outliers in some of the dependent variables, the analyses were again repeated with Winsorised data at the .05 level (i.e. outliers were replaced with the 5th or 95th percentile values, depending on the tail of the distribution).

Tests on both the non-Winsorised and Winsorised data revealed a significant effect of *type of resource* on only one of the dependent variables, *accuracy* (problems) (Non-Winsorised: $\chi 2(1) = 4.481$, $p = .034$; Winsorised: $\chi 2(1) = 4.701$, $p = .030$). There was no significant effect of *type of resource* on any of the other dependent variables, including *productivity* (problems) (Non-Winsorised: $\chi 2(1) = 0.845$, $p = .358$; Winsorised: $\chi 2(1) = 0.845$, $p = .348$), *productivity* (solutions) (Non-Winsorised: $\chi 2(1) = 0.506$, $p = .477$; Winsorised: $\chi 2(1) = 0.506$, $p = .477$), *accuracy* (solutions) (Non-Winsorised: $\chi 2(1) = 3.361$, $p = .067$; Winsorised: $\chi 2(1) = 3.361$, $p = .067$), and *confidence* (Non-Winsorised: $\chi 2(1) = 0.604$, $p = .437$; Winsorised: $\chi 2(1) = 0.620$, $p = .431$). These results indicate that participants' accuracy in correctly identifying problems in the Pure Gym homepage was greater when using WebAIR than when using WCAG 2.0. For all other variables, the accessibility information resource used had no effect on the participants' performance.

A series of Chi Square analyses compared participants' productivity and accuracy in the eight most frequently attempted questions in the Pure Gym homepage accessibility evaluation task. The small number of participants attempting some of the questions violated the cell size assumption of the Chi Square test. Again, where the expected frequencies were less than five, Fisher's exact test was calculated instead. Results of the analyses are presented in Table 7.7 and Table 7.8, respectively. The analyses revealed no significant differences in the *productivity* measures between the two conditions for either the problem or solution aspect of any of the questions. These results indicate that participants were no more or no less productive in any of the eight most frequently attempted questions using either resource. However, the analysis revealed a significant difference in the *accuracy* measures between the two conditions for the problem aspect of Question 6 and the solution aspect of Question 7. For Question 6, participants using WebAIR were significantly more accurate in correctly identifying the problem than participants using WCAG 2.0. For Question 7, participants using WCAG 2.0 were significantly more accurate in correctly identifying the solution than participants using WebAIR. For the problem and solution aspect of all other questions, there were no significant differences in the accuracy measures between the two conditions. These results indicate that, with the exception of the problem aspect of Question 6 and the solution aspect of Question 7, participants were no more or no less accurate in any of the eight most frequently attempted questions using either resource.

Descriptive statistics of the participants' self-rating of *confidence* in their responses to the eight most frequently attempted questions in the Pure Gym homepage accessibility evaluation task are presented in Table 7.9. A further series of one-way non-parametric Kruskal-Wallis tests investigated the effect of *type of resource* (WebAIR or WCAG 2.0) on the mean self-rating of *confidence* for each individual question. The tests revealed no significant effect of *type of resource* on participants' confidence for any of the questions. These results indicate that participants were no more or no less confident in their responses to any of any of the eight most frequently attempted questions using either resource.

Table 7.7: Frequencies and Chi Square test/Fisher's Exact test results of participants' productivity in the eight most frequently attempted questions in the Pure Gym homepage accessibility evaluation task

			WebAIR	*WCAG 2.0*	*Total*					
		N	*Attempted*	*Not attempted*	*Attempted*	*Not attempted*	*Attempted*	*Not attempted*	*χ(1)*	*P-Value*
Q1	***Problem***	48	19	3	23	3	42	6	N/A	1.000
	Solution	48	17	5	19	7	36	12	0.112	0.738
Q2	***Problem***	48	22	0	25	1	47	1	N/A	1.000
	Solution	48	21	1	22	4	43	5	N/A	0.357
Q3	***Problem***	48	22	0	21	5	43	5	N/A	0.054
	Solution	48	15	7	14	12	29	19	1.024	0.312
Q4	***Problem***	48	17	5	17	9	34	14	0.815	0.367
	Solution	48	12	10	11	15	23	25	0.715	0.398
Q5	***Problem***	48	16	6	21	5	37	11	0.436	0.509
	Solution	48	12	10	15	11	27	21	0.048	0.827
Q6	***Problem***	48	14	8	12	14	26	22	1.467	0.226
	Solution	48	5	17	10	16	15	33	1.373	0.241
Q7	***Problem***	48	10	12	13	13	23	25	0.099	0.753
	Solution	48	6	16	8	18	14	34	0.071	0.791
Q8	***Problem***	48	11	11	11	15	22	26	0.284	0.594
	Solution	48	9	13	8	18	17	31	0.536	0.464

Table 7.8: Frequencies and Chi Square test/Fisher's Exact test results of participants' accuracy in the seven most frequently attempted questions in the Pure Gym homepage accessibility evaluation task

			WebAIR		***WCAG 2.0***	***Total***				
		N	***Correct***	***Incorrect***	***Correct***	***Incorrect***	***Correct***	***Incorrect***	***χ(1)***	***P Value***
Q1	***Problem***	42	15	4	19	4	34	8	N/A	1.000
	Solution	36	10	7	11	8	21	15	0.003	0.955
Q2	***Problem***	47	5	17	3	22	8	39	N/A	0.446
	Solution	43	3	18	1	21	4	39	N/A	0.345
Q3	***Problem***	43	6	16	2	19	8	35	N/A	0.240
	Solution	29	4	11	1	13	5	24	N/A	0.330
Q4	***Problem***	34	12	5	9	8	21	13	1.121	0.290
	Solution	23	6	6	4	7	10	13	N/A	0.680
Q5	***Problem***	37	12	4	13	8	25	12	0.711	0.399
	Solution	27	10	2	10	5	20	7	N/A	0.408
Q6	***Problem***	26	9	5	3	9	12	14	4.013	**0.045**
	Solution	15	0	5	0	10	0	15	N/A	N/A
Q7	***Problem***	23	5	5	8	5	13	10	N/A	0.685
	Solution	14	2	4	8	0	10	4	N/A	**0.015**
Q8	***Problem***	22	10	1	8	3	18	4	N/A	0.586
	Solution	17	9	0	7	1	16	1	N/A	0.471

Table 7.9: Descriptive statistics relating to participants' confidence in their responses to the eight most frequently attempted questions in the Pure Gym homepage accessibility evaluation task

	WebAIR				*WCAG 2.0*					
	N	*Mean*	*Std. Dev.*	*Min.*	*Max.*	*N*	*Mean*	*Std. Dev.*	*Min.*	*Max.*
Q1	19	3.32	1.157	1	5	22	3.09	1.109	1	5
Q2	22	4.32	0.716	3	5	25	4.40	0.866	2	5
Q3	21	3.67	1.017	1	5	20	3.75	0.910	1	5
Q4	14	3.71	1.204	1	5	16	3.00	1.033	1	5
Q5	14	3.86	1.231	1	5	17	3.47	1.328	1	5
Q6	13	3.77	0.832	2	5	12	3.17	1.115	1	5
Q7	9	3.00	0.707	2	4	12	2.92	1.311	1	5
Q8	10	3.90	1.101	2	5	11	3.45	1.214	1	5

7.4. Discussion

The aim of this study was to evaluate the effectiveness of both WebAIR and WCAG 2.0 in supporting novice web developers to confidently identify accessibility problems and solutions in existing websites. The study provided the first opportunity to evaluate WebAIR following its substantial revision in response to the previous study. More specifically, the study provided the opportunity to evaluate the effectiveness of WebAIR in supporting web developers who are relatively new to web accessibility.

Forty-eight novice web developers alternatively used both WebAIR and WCAG 2.0 to identify accessibility problems and solutions in the homepages of two publicly-available websites: Manchester United football club and the Pure Gym gym operator. As time available was limited, instead of conducting full web accessibility evaluations of the two webpages, participants answered a series of questions that would guide them towards identifying potential accessibility problems in the evaluation webpages.

Participants rated WebAIR significantly more highly than WCAG 2.0 across a number of characteristics, including usefulness, ease of use, navigability, understandability, clarity of organisation, and likelihood of using the resource. Although participants rated WCAG 2.0 significantly more highly than WebAIR on amount of information, there was no significant difference between the two on ratings of either completeness or the number of items to test. These results provide subjective measures of the effectiveness of the two resources, indicating that, overall, participants prefer WebAIR. However, the apparent effectiveness of WebAIR was not reflected in the objective measures of the accessibility evaluation tasks.

In the evaluation of both the Manchester United and Pure Gym homepages, there was no significant multivariate effect of accessibility information resource on participants' productivity, accuracy or confidence in identifying either problems or solutions. Overall, in both evaluations, participants were no more productive, accurate or confident in their responses using either WebAIR or WCAG 2.0. Though follow-up univariate tests confirmed this result for the Manchester United homepage, they indicated that participants' *accuracy* in correctly identifying problems in the Pure Gym homepage was significantly greater when using WebAIR than when using WCAG 2.0.

A question-by-question analysis of the most commonly attempted questions in both evaluations served to support and clarify these findings. There was no significant effect of accessibility information resource either on participants' productivity or accuracy for any of the seven most frequently attempted questions in the Manchester United homepage evaluation. This supports the overall finding that the accessibility information resource used had little bearing upon participants' performance for this webpage evaluation. For two of the questions, however, participants were significantly more confident using WebAIR than WCAG.

For the Pure Gym webpage evaluation, the accessibility information resource used made no difference to participants' confidence to any of the eight most frequently attempted questions. However, it had an effect on participants' responses to two of the questions. Participants using WebAIR were significantly more accurate in correctly identifying the problem in Question 6 than participants using WCAG 2.0. Conversely, participants using WCAG 2.0 were significantly more accurate in correctly identifying the solution in Question 7 than participants using WebAIR.

These unusual results could be explained by the ordering of the items in WebAIR. Answers to Question 6—*Could people with disabilities experience any problems in knowing what goes where in the 'Member Login' form?*—can be found in the very first question in WebAIR: *Do all inputs and controls have text labels that accurately describe what they are for?* The ordering of the items in WebAIR therefore may have resulted in an unintentional primacy effect, making it easier for participants to identify the problem for that particular question. Such an effect does not, however, account for participants' greater accuracy in correcting identifying the solution to Question 7 using WCAG 2.0. Answers to Question 7—*Have the headings on the webpage been correctly implemented for accessibility (e.g. using a logical hierarchy of headings)?* —can be found in Success criterion 1.3.1 Info and Relationships, which is several sections into the resource. Also, participants were no more accurate using WCAG 2.0 for questions relating to text alternatives (e.g. Question 2—*Could the image of an ice cube used in the 'freeze your membership' section of the webpage cause any problems for people with disabilities?),* which is the very first topic addressed in WCAG 2.0 Success criterion 1.1.1.

The results indicate that despite participants rating the resource significantly more highly across a range of characteristics, WebAIR is no more effective (but, equally, no less effective) than WCAG 2.0 in supporting novice web developers to confidently identify accessibility problems and solutions in existing websites. Although hopefully the design decisions made in the development of WebAIR, relating to its language and terminology, organisation and structure, and volume and comprehensiveness, may have contributed to participants' positive perceptions of the resource, they appear to have had little impact on participants' effectiveness in using the resource to evaluate the accessibility of websites.

This disappointing outcome, however, may have had more to do with the nature of the evaluation and how it encouraged participants to use the accessibility information resources. Taking into consideration both the students' inexperience and the limited amount of time available in which to conduct the study, I made the decision to have participants answer a series of questions that would guide them towards identifying potential accessibility problems in the evaluation webpages. Instead of identifying accessibility problems and solutions without any guidance, participants were instead prompted to consider the accessibility of particular webpage features. This represented a necessary trade-off between evaluating the accessibility information resources effectively and providing a useful learning experience for the students involved.

While this approach provided the necessary guidance to inexperienced participants, it may have also encouraged them to use the resources to support and verify, rather than inform, their own judgements. The individual evaluation questions directed participants to particular areas of the evaluation webpages, but instead of provoking them to find the answer in the resources, the purposely-ambiguous phrasing of the questions (e.g. *Could X cause any problems for people with disabilities?*) appeared to encourage them to make their own judgements. Participants were also aware that a number of 'red herring' questions had been included for which an accessibility problem did not exist in the evaluation webpages. As a result, they tended to diagnose accessibility problems from their own knowledge before looking to the resources to support or verify those diagnoses. This strategy was disclosed in one participant's response to a question about headings. Asked whether a lack of headings could be a problem for people with disabilities, the participant responded: "It probably would, but I can't find a WebAIR guideline to support this" [D3]. Another participant incorrectly interpreted a question about decorative images, commenting: "I have no idea where to find what I'm looking for with WCAG" [C17].

While this strategy may be successful for professional web developers with some knowledge of web accessibility, the relative inexperience of the novice web developers in this study often led to them either incorrectly diagnosing problems or dismissing them completely. They then struggled to find the relevant section of the resources to resolve the problem. For example, one of the questions about the Manchester United webpage referred to missing link text, which would cause screen readers to not announce the link properly. One participant incorrectly assumed the problem related to colour contrast and cited sections of WebAIR relating to colour contrast in their response. Similarly, a question referring to inadequate text alternatives on the Pure Gym webpage was incorrectly assumed to be about keyboard navigation by one participant, who cited sections of WCAG 2.0 relating to tab order in their response. Many participants neglected to answer the solution aspect of the questions, presumably because they had incorrectly diagnosed the problem and were unable to find the relevant section of the resources to support them.

Using the resources to support and verify accessibility problems relies on participants having correctly diagnosed the problems in the first place, which is not something the resources can help with. This observation highlights one of the drawbacks of such resources for web developers less familiar with web accessibility. However well-organised or clearly presented such resources may be, without a basic grounding in web accessibility that allows them to correctly diagnose accessibility problems, novice web developers will struggle to locate the relevant guidance in the resources and to apply it to their websites. Further study is necessary to explore how web developers actually use the accessibility information resources and to investigate whether professional web developers, who may be more familiar with the subject, are more successful in their approach.

Aside from the limitation posed by the use of questions rather than an open evaluation method, the study had other limitations that need to be considered.

The student participants were not offered any money or course credit to take part in the study. Instead, the study was promoted as a relevant learning experience from which the students could benefit and be able to use in their module assessment. Despite this potential educational incentive, the students were not particularly motivated to take part. This was noticeable from the often terse or flippant nature of their responses.

Several participants objected to the notion of conducting an experiment during their practical session and, on this basis, one participant did not consent to take part. Because of participants' potential lack of motivation, they may not have put much effort into the exercise, thus reducing the validity of the results.

Another limitation was the use of live websites in the accessibility evaluation tasks. Although this provided more ecologically valid stimuli, by giving participants the opportunity to evaluate genuine websites with genuine accessibility issues, it meant that the websites could not be controlled. This danger was exemplified by one of the evaluation websites (Manchester United) being updated moments before the study was due to commence. Thankfully the update was fairly superficial, and the affected study question could be adjusted but had it been a more significant update, it would have severely compromised the study.

Although participants were asked to complete as many of the study questions as possible in the time allowed, there was no requirement for them to complete them sequentially or in any particular order. When it came to assessing the participants' responses, it was difficult to establish whether missing responses were due to participants not having attempted or not having reached the question or because they had struggled to provide a response. Similarly, as mentioned above, it was difficult to establish whether participants had actually used the resources in answering the study questions.

A further limitation related to the requirement for participants to type their responses to the questions into a spreadsheet. With the already limited amount of time, the need to formulate written responses (rather than, for example, describing their responses out loud) may have added an extra burden and limited their productivity in the accessibility evaluation task.

7.5. Conclusions

The outcomes of this study were disappointing. It provided the first opportunity to evaluate the effectiveness of the revised version of WebAIR under controlled experimental conditions. Yet despite participants rating the resource significantly more highly across a range of characteristics, WebAIR appeared to be no more effective (but, equally, no less effective) than WCAG 2.0 in supporting novice web developers to confidently identify accessibility problems and solutions in existing websites. It is tempting to attribute this disappointing outcome to the numerous limitations of the study. Certainly, the laboratory-based nature of the experimental and the limited amount of time that participants were exposed to the resources may have weakened the internal and external validity of the experiment, as could the inadequate incentives for participants to take part. However, neither time nor enthusiasm are guaranteed attributes for any web developer undertaking web accessibility evaluations. A more productive conclusion to draw from this study is the possibility that the design of WebAIR is less effective for web developers who are less familiar with web accessibility. Any advantage that WebAIR's design might have over that of existing accessibility information resources—and the participants' ratings seems to suggest it at least has face value—appears to be negated by the need to have at least some knowledge of web accessibility to be able to use the resource effectively. This is an important consideration in light of my attempt to improve access to web accessibility information by web developers who, as the previous studies have demonstrated, may have little knowledge of the subject. The evaluation of WebAIR continues in the following chapter, which addresses some of the limitations of the current study by evaluating WebAIR under more ecologically valid experimental conditions.

Chapter 8. Further evaluation of WebAIR's effectiveness in supporting web accessibility evaluation

8.1. Introduction

The evaluation of WebAIR presented in the previous chapter assessed the effectiveness of the resource in supporting novice web developers to confidently identify accessibility problems and solutions in existing websites. The outcome of the evaluation was disappointing and raised the possibility that, despite being well received by novice web developers, WebAIR may be less effective for newcomers to accessibility.

The laboratory-based nature of the previous study, coupled with the participants' limited exposure to the resource and the participants' lack of motivation, threatened both the internal and external validity of the experiment. The use of live websites, though authentic, also proved problematic and impossible to control. To adequately investigate the effectiveness of WebAIR in supporting web developers to create and evaluate accessible websites, a rigorous, yet ecologically valid experimental design was necessary.

This chapter presents a "real-world" evaluation of WebAIR with 32 professional web developers. As in the previous chapters, despite one of the aims of WebAIR being to further web developers' knowledge and understanding of web accessibility, this aspect of the resource could not adequately be evaluated in this study. Although the amount of time that participants were exposed to the resource was further increased to 35 minutes, it was still not sufficient to observe any substantial changes in knowledge or understanding. Instead, the study objective was to evaluate the ease of use and effectiveness of the resource in supporting professional web developers to identify accessibility problems and solutions in a purpose-built website. The study focuses in particular on how the web developers actually use the resources to conduct accessibility evaluations.

8.2. Method

8.2.1. Design

To investigate the effectiveness of WebAIR in supporting professional web developers to evaluate the accessibility of websites, a between-participants with 32 professional web developers study was conducted. Participant used either WebAIR or WCAG 2.0 to identify and resolve accessibility problems in the homepage of Eat My Goal![68] a purpose-built football news website designed to incorporate a range of accessibility problems (see Figure 8.1).

The independent variable was the type of resource (WebAIR and WCAG 2.0) used by the web developers. The dependent variable was the number of problems correctly identified by each web developer. Participants also used a concurrent think aloud protocol to describe what they were doing and thinking. The participants' reactions to the accessibility information resources were elicited using rating items.

To allow for greater flexibility in recruiting participants, and also to provide the opportunity to compare remote and local methods, half of the evaluations were conducted in person, at a location convenient to each participant, and half remotely, using the video conferencing software, GoToMeeting[69].

The methodology was piloted on a web developer who is a friend of the author, with the procedure refined in light of that experience.

8.2.2. Participants

Thirty-two professional web developers took part in the study. Four participants were female, the rest were male. The age of participants ranged from 20 to 44 years (mean: 31.0). One participant was from Australia, another was from The Netherlands, the rest were from the UK. Participants were recruited using opportunistic sampling from web developer mailing lists, online forums, social media, conference flyers and personal contacts in the industry. The participants had between 1 and 20 years of experience of web development (mean: 10 years). Eight participants worked for large enterprises (250+ employees), eighteen worked for SMEs (< 250 employees) and six participants

---

[68] https://www.cs.york.ac.uk/hci/dswallow/eatmygoal/

[69] https://www.gotomeeting.com/

were self-employed freelancers. Given the verbal nature of the tasks, participants were screened for their spoken English proficiency. Participants were offered £20 worth of Amazon vouchers in compensation for their time and effort.

Participants were required to be familiar with developing websites but were not required to be familiar with web accessibility. Prior to the evaluation, participants reported being fairly familiar with web accessibility (mean: 3.9; SD: 0.71) and moderately familiar with WCAG (mean: 3.1; SD: 1.1) (both scales: 1 = Not at all familiar, 5 = Very familiar).

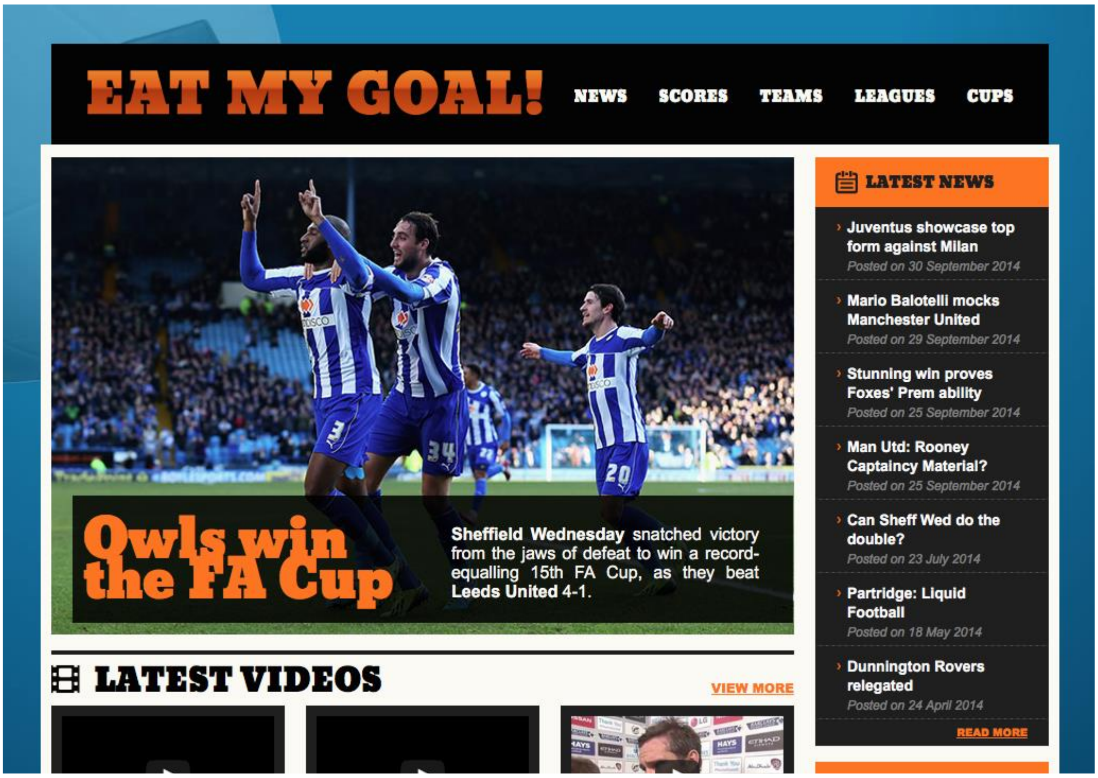


Figure 8.1: Eat My Goal! football news homepage

8.2.3. Materials

In both the face-to-face and remote evaluation sessions, participants used their own computer with Internet access. The GoToMeeting software, which offers screen and audio recording, was used in *both* face-to-face and remote evaluation sessions to video and audio record the evaluation.

At the start of each session, participants in both the face-to-face and remote evaluation sessions viewed a webpage containing information about the study and completed an online informed consent form (see Appendix A). At the end of each session, participants completed a Likert rating scale relating to their experience of using the resource, and also a short demographic questionnaire.

Participants conducted a website accessibility evaluation using either the complete version of WebAIR or the complete set of WCAG 2.0 guidelines. To eliminate the variability of working with a live website, and also to control for the number of accessibility problems that participants might identify, participants evaluated the homepage of a fictional football news website called Eat My Goal!, created using HTML, CSS and JavaScript and purposely designed it to include a range of accessibility problems.

The study webpage presented 45 distinct accessibility problems for participants to identify. Table 8.1 describes each problem, its location within the Eat My Goal! website, and its priority (indicated by the WCAG 2.0 conformance level it represents).

Table 8.1: Accessibility problems in the Eat My Goal! website

ID	Description	Location	WCAG 2.0 Conf. Level
1	No indication that the list of links is a navigation menu	Header	A
2	No indication of the current location within the navigation menu	Header	AAA
3	No skip links to main content or specific sections of webpage	Header	A
4	Page title not very descriptive	Header	A
5	Logo image link lacks text alternative	Header	A
6	Insufficient colour contrast (orange text/black background)	Header	AA
7	Lead story image lacks text alternative	Lead story	A
8	Inappropriate markup used to define bold text	Lead story	A
9	Lead story text is justified (aligned to both left and right margins)	Lead story	AAA
10	Video content not keyboard accessible	Latest videos	A
11	Inappropriate markup used to define paragraphs	Latest videos	A
12	Video content lacks text alternative	Latest videos	A
13	Video content lacks captions and/or transcripts	Latest videos	A
14	Video content lacks audio description	Latest videos	A
15	Video content lacks sign language interpretation	Latest videos	AAA
16	Ambiguous and repetitive “View More” links	Latest videos	A
17	Insufficient colour contrast (blue text/black background)	Latest videos	AA
18	Videos trap keyboard focus	Latest videos	A

ID	*Description*	*Location*	*WCAG 2.0 Conf. Level*
19	Adjacent image and text links not combined	Featured articles	A
20	Image text alternatives not very descriptive	Featured articles	A
21	Insufficient colour contrast (orange text/white background)	Featured articles	AA
22	Headings not very descriptive	Featured articles	AA
23	Insufficient line spacing on paragraph text	Featured articles	AAA
24	Inappropriate markup used to define emphasised text	Latest news	A
25	Inappropriate markup used to define headings	Latest news	A
26	Insufficient colour contrast (grey text/black background)	Fixtures	AA
27	Presentation of text not controlled using CSS	Fixtures	A
28	Text labels not associated with form controls	Newsletter	A
29	Required fields indicated by colour alone	Newsletter	A
30	No visible focus indicator on form fields	Newsletter	AA
31	No instructions on how to complete form	Newsletter	A
32	No feedback when form is submitted successfully	Newsletter	A
33	Feedback does not indicate how to resolve errors	Newsletter	A
34	Links open in new window without warning	Footer	A
35	List element not used to group related items	Footer	A
36	Links indicated by colour alone	Footer	A
37	Link text does not describe purpose of link	Footer	A
38	Insufficient colour contrast (blue text/black background)	Footer	AA
39	Insufficient colour contrast (grey icon/black background)	Footer	AA
40	Illogical heading hierarchy	General	A
41	User-specified colour schemes are overwritten	General	AA
42	Text cannot be resized to 200% without loss of content	General	AA
43	Language of webpage not specified	General	A
44	Audio plays automatically for more than 3 seconds	General	A
45	No means of pausing or stopping audio	General	A

8.2.4. Procedure

Face-to-face evaluations were conducted at a location convenient to participants. In many cases, this was at the participant's workplace, but also cafés, restaurants, and at participants' homes. The remote evaluations were conducted from either the author's home or workplace using GoToMeeting. Participants in the remote evaluations did not specify their location.

The same procedure was followed for both face-to-face and remote evaluation sessions. Each session began with a brief introduction and explanation of the purpose of the study. Participants then had the opportunity to read the study information webpage, before completing the first part of the online consent form. Participants were then given a brief demonstration of how to use the accessibility information resource to address an accessibility problem (but not one of the accessibility problems that they might encounter in the experimental website). Participants then rated their familiarity both with web accessibility and with WCAG on two 5-point items (1 = Not at all familiar; 5 = Very familiar).

Participants then had five minutes to familiarise themselves with either WebAIR or WCAG 2.0 and to ask any questions about the study.

When participants were ready to begin the evaluation, they were given access to the experimental webpage, Eat My Goal!, and asked to identify as many problems as possible within 35 minutes using the resource they had been allocated.

At this stage, participants were assured that because there were plenty of accessibility problems in the experimental webpage and because both resources are fairly large, they should not feel they have to cover everything. It was emphasised that the purpose of the study was to investigate the effectiveness of the accessibility information resources and that participants should not worry if they did not feel very confident about the evaluation. Participants were also told that it was up to them how they approached the task and what techniques they used. For example, some problems require participants to visually inspect or use the webpage (e.g. if they need to determine how something is presented to the user, such as an image or the wording of text), whereas others require participants to view the source code of the webpages (e.g. if they need to determine whether the appropriate code has been used in an accessible way). Though participants were permitted to use external tools and applications (such as code validators), they were asked to refrain from using automated accessibility evaluation tools.

Irrespective of how participants approached the task, they were asked to describe in as much detail as possible what they are doing and thinking, using a concurrent think aloud protocol. They were asked to describe the problems they identified and to describe their use of the resources. In particular, they were asked to mention if they felt there is something missing from the resource or if they experienced any problems with it.

As participants undertook the task, the list of accessibility problems included in the experimental webpage was reviewed, with the problems that participants found being checked off, along with the time they found them. Any additional problems or comments that participants mentioned were noted. In addition to this, the way in which participants approached the task and any strategies they used were also noted.

Once the task was complete, participants were given a list of the accessibility problems in the experimental webpage and indicated whether they had identified them correctly. They were also asked to comment on any problems that they had failed to identify or had dismissed as not being a problem.

Participants then completed a short questionnaire eliciting their impressions of the resource. This was the same questionnaire as used in the previous two studies (see Section 6.3.3).

Finally, participants completed a brief demographics questionnaire and the second part of the online consent form, in which they provide their consent to submit their responses.

Once the evaluation session was complete and the participants had been debriefed, they were thanked for their time and asked whether they would be interested in taking part in further research. The evaluation sessions lasted approximately 60 minutes each.

8.2.5. Data preparation

In addition to the notes taken during each evaluation session, the video recordings were later reviewed to confirm and elaborate upon the initial observations. Contrary to the previous studies, the video recordings were not fully transcribed. Not only had verbatim transcripts proven unnecessary in the previous studies, they had also been extremely time-consuming to produce. Instead, the notes from both the evaluation sessions and later review were imported into QSR NVivo and analysed using content analysis (Hsieh and Shannon, 2005), as in the previous studies (see Sections 5.3.5 and 6.3.5). The unit of analysis in this case was each comment/observation, to which multiple codes could be applied. Sample notes from four participants can be found in Appendix G.

The comments/observations were categorised according to the type of issue on which the participant had commented (e.g. clarity of terminology, repetition of content, expected information not found etc.) or the strategy/behaviour that had been observed (e.g. participant focuses on automated accessibility issues, participant works from the code, participant applies existing knowledge etc.). From this open coding approach, an initial set of 27 issue codes and 13 strategy/behaviour codes were derived. Related issue codes were then grouped into the same 18 subcategories and 3 broad categories that had been defined in the earlier study (a list of codes and categories can be found in Appendix G).

As in the two previous studies, the analysis was validated by an independent researcher. Again, inter-coder reliability was not undertaken but the researcher coded the notes on four of the participants (approximately 10% of the data) using the same codes developed during my analysis. A degree of consistency between the researcher's and my application of the codes would then serve to corroborate, if not confirm, the results. As in the previous studies, any disagreements between the two coding attempts (of which there were very few) were resolved through discussion.

8.3. Results

Figure 8.2, below, shows the mean ratings of various characteristics of WebAIR and WCAG 2.0. Descriptive statistics of the mean ratings are presented in Table 8.2.

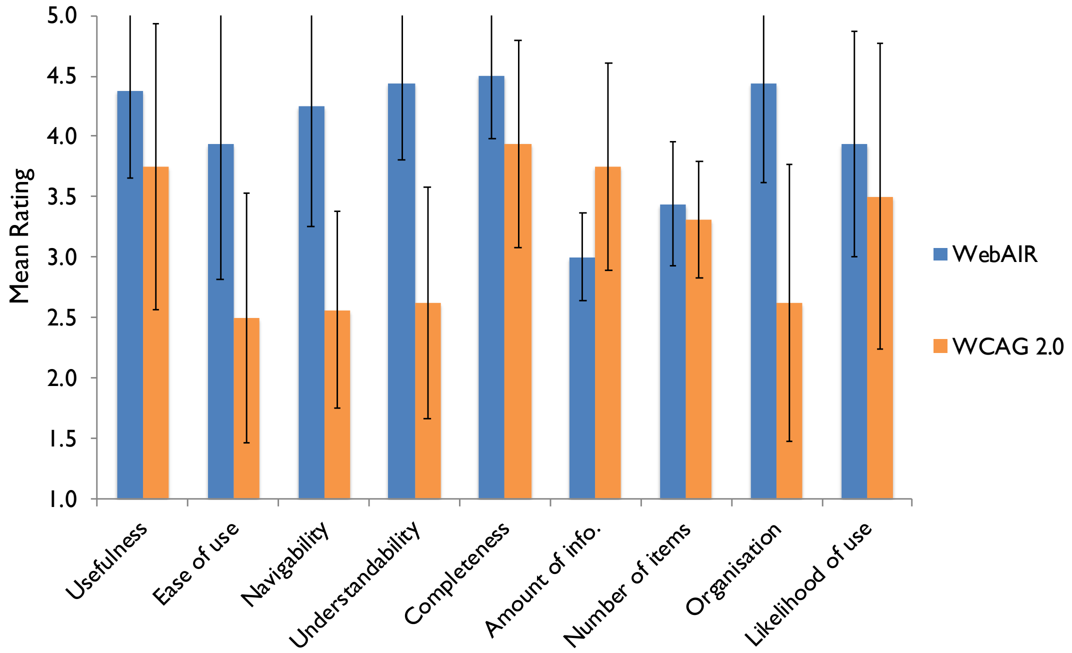


Figure 8.2: Mean ratings of WebAIR and WCAG 2.0 on a range of characteristics

Error bars: standard deviation

A series of one-way ANOVAs investigated the effect of *type of resource* (WebAIR or WCAG 2.0) on the ratings of the characteristics. Again, while conducting multiple tests may inflate the Type 1 error rate, it nevertheless provides useful comparisons between the different resources.

The tests showed significant effects of *type of resource* on all of the ratings except *usefulness*, $F(1, 30) = 3.26$, $p = .081$, $\eta^2_p = .10$, *number of items to test*, $F(1, 30) = 0.51$, $p = .481$, $\eta^2_p = .02$, and *likelihood of using*, $F(1, 30) = 1.24$, $p = .274$, $\eta^2_p = .04$. WebAIR was rated significantly higher on *ease of use*, $F(1, 30) = 14.20$, $p = .001$, $\eta^2_p = .32$, *navigability*, $F(1, 30) = 27.41$, $p < .001$, $\eta^2_p = .48$, *understandability*, $F(1, 30) = 40.05$, $p < .001$, $\eta^2_p = .57$, *completeness*, $F(1, 30) = 5.08$, $p = .032$, $\eta^2_p = .15$, *organisation*, $F(1, 30) = 26.56$, $p < .001$, $\eta^2_p = .47$. WCAG 2.0 was rated significantly higher on *amount of information*, $F(1, 30) = 10.39$, $p = .003$, $\eta^2_p = .26$.

Table 8.2: Descriptive statistics of mean ratings of WebAIR and WCAG 2.0 on a range of characteristics

Characteristic		***WebAIR***	***WCAG 2.0***
Usefulness	***Mean***	4.4	3.8
	SD	0.72	1.18
Ease of use	***Mean***	3.9	2.5
	SD	1.12	1.03
Navigability	***Mean***	4.3	2.6
	SD	1.00	0.81
Understandability	***Mean***	4.4	2.6
	SD	0.63	0.96
Completeness	***Mean***	4.5	3.9
	SD	0.52	0.85
Amount of information	***Mean***	3.0	3.75
	SD	0.37	0.86
Number of items to test	***Mean***	3.4	3.3
	SD	0.51	0.48
Organisation	***Mean***	4.4	2.6
	SD	0.81	1.15
Likelihood of using	***Mean***	3.9	3.5
	SD	0.93	1.27

8.3.1. Eat My Goal! homepage accessibility evaluation task

Overall, participants identified between 4 and 30 problems in each evaluation session (mean: 15.5 problems; SD: 6.02). The mean represents just over a third (34%) of the 45 problems in the webpage. Participants using WebAIR identified between 8 and 30 problems in each evaluation session (mean: 17.8 problems; SD: 6.19). The mean represents 40% of the 45 problems in the webpage. Participants using WCAG 2.0 identified between 4 and 22 problems in each evaluation session (mean: 13.2; SD: 5.01). The mean represents 29% of the 45 problems in the webpage.

An independent-samples t-test compared the number of problems identified by participants using WebAIR and WCAG 2.0. This showed a significant effect of *type of resource*, $t(30) = 2.32$, $p < .05$; $d = 0.82$. The effect size for this analysis ($d = 0.82$) was found to exceed Cohen's (1988) convention for a large effect ($d = 0.80$). This result indicates that participants identified a significantly greater number of problems using

WebAIR (mean: 17.8; SD: 6.19) to evaluate the accessibility of the webpage than using WCAG 2.0 (mean: 13.2; SD: 5.01).

Each problem was identified by at least one participant using either resource. Twenty-six participants identified the most frequently identified problem; only one identified the least frequently identified problem. Table 8.3 describes the most frequently identified problems by participants both overall and broken down by the resource used.

Table 8.3: Most frequently identified problems overall

		Number (percentage) of participants		
ID	*Description*	*Overall*	*Using WebAIR*	*Using WCAG 2.0*
5	Logo image link lacks text alternative	26 (81%)	11 (69%)	15 (94%)
20	Image text alternatives not very descriptive	24 (75%)	13 (81%)	11 (69%)
4	Page title not very descriptive	23 (72%)	11 (69%)	12 (75%)
29	Required fields indicated by colour alone	23 (72%)	16 (100%)	7 (44%)
28	Text labels not associated with form controls	21 (66%)	13 (81%)	8 (50%)

Table 8.4 and Table 8.5 describe the most frequently identified problems by participants using WebAIR and WCAG 2.0, respectively.

Table 8.4: Most frequently identified problems by participants using WebAIR

ID	*Description*	*Number (percentage) of participants*
29	Required fields indicated by colour alone	16 (100%)
20	Image text alternatives not very descriptive	13 (81%)
28	Text labels not associated with form controls	13 (81%)
32	No feedback when form is submitted successfully	13 (81%)
5	Logo image link lacks text alternative	11 (69%)
4	Page title not very descriptive	11 (69%)
30	No visible focus indicator on form fields	11 (69%)

Table 8.5: Most frequently identified problems by participants using WCAG 2.0

ID	*Description*	*Number (percentage) of participants*
5	Logo image link lacks text alternative	15 (94%)
4	Page title not very descriptive	12 (75%)
20	Image text alternatives not very descriptive	11 (69%)
44	Audio plays automatically for more than 3 seconds	11 (69%)
40	Illogical heading hierarchy	10 (63%)
45	No means of pausing or stopping audio	10 (63%)

The most frequently identified problems by participants using WebAIR related to forms and alternative text descriptions of images and links. The most frequently identified problems by participants using WCAG 2.0 related to alternative text descriptions of images and links as well as whole page issues, such as the webpage title, audio and heading structure.

Table 8.6 describes the least frequently identified problems by participants both overall and broken down by the resource used.

Table 8.6: Least frequently identified accessibility problems

		Number (percentage) of participants		
ID	***Description***	***Overall***	***Using WebAIR***	***Using WCAG 2.0***
24	Inappropriate markup used to define emphasised text	1 (3%)	1 (6%)	0 (0%)
35	List element not used to group related items	2 (6%)	2 (13%)	0 (0%)
6	Insufficient colour contrast (orange text/black background)	2 (6%)	1 (6%)	1 (6%)
10	Video content not keyboard accessible	3 (9%)	2 (13%)	1 (6%)
14	Video content lacks audio description	4 (13%)	2 (13%)	2 (13%)
27	Presentation of text not controlled using CSS	4 (13%)	3 (19%)	1 (6%)

Table 8.7 and Table 8.8 describe the least frequently identified problems by participants using WebAIR and WCAG 2.0, respectively.

Table 8.7: Least frequently identified problems by participants using WebAIR

ID	***Description***	***Number (percentage) of participants***
24	Inappropriate markup used to define emphasised text	1 (6%)
6	Insufficient colour contrast (orange text/black background)	1 (6%)
15	Video content lacks sign language interpretation	1 (6%)
2	No indication of the current location within the navigation menu	1 (6%)
35	List element not used to group related items	2 (13%)
10	Video content not keyboard accessible	2 (13%)
14	Video content lacks audio description	2 (13%)
22	Headings not very descriptive	2 (13%)

Table 8.8: Least frequently identified problems by participants using WCAG 2.0

ID	*Description*	*Number (percentage) of participants*
24	Inappropriate markup used to define emphasised text	0 (0%)
35	List element not used to group related items	0 (0%)
42	Text cannot be resized to 200% without loss of content	0 (0%)
6	Insufficient colour contrast (orange text/black background)	1 (6%)
10	Video content not keyboard accessible	1 (6%)
27	Presentation of text not controlled using CSS	1 (6%)
37	Link text does not describe purpose of link	1 (6%)
21	Insufficient colour contrast (orange text/white background)	1 (6%)
34	Links open in new window without warning	1 (6%)
11	Inappropriate markup used to define paragraphs	1 (6%)

The least frequently identified problems by participants using WebAIR related to whole page issues, such as inappropriate use of HTML and CSS and accessibility of the video content. The least frequently identified problems by participants using WCAG 2.0 also related to whole page issues, such as inappropriate use of HTML and CSS, webpage resizing and colour contrast.

8.3.2. Observations of accessibility information resource usage

Based on the participants' self-reported usage data and my own observations, I investigated how they approached the website accessibility evaluation task and whether the two resources were effective in supporting their behaviour.

Irrespective of the resource they were using, participants employed noticeably different strategies in completing the task.

Twelve participants began from the evaluation website, inspecting both the content and the underlying code for potential problems, before actively searching and consulting the resource for solutions. The structure and organisation of the website thus provided an order or sequence for participants to follow. Participants who took this *searching and consulting* approach tended to use the resource as a reference tool for looking up and learning about specific accessibility issues. Eight of the 16 participants who used WCAG 2.0, and 4 of the 16 who used WebAIR, took this approach.

The other twenty participants began from the resource, following the accessibility information and guidance it contains before then applying it to the evaluation website. The structure and organisation of the resource thus provided an order or sequence for participants to follow. Participants who took this *following and applying* approach tended to treat the resource as a prompt or checklist, against which they could evaluate the website. Eight of the 16 participants who were allocated WCAG 2.0, and 12 of the 16 who were allocated WebAIR, took this approach.

8.3.2.1. Searching and consulting using WCAG 2.0

The eight participants who took a *searching and consulting* approach to the website accessibility evaluation task using WCAG 2.0 identified between 4 and 22 accessibility problems (mean: 14.0 problems; SD: 6.93).

WCAG 2.0 presents an exhaustive collection of accessibility recommendations and practical techniques but its structure and organisation prevented some participants from using it effectively. Having identified potential accessibility issues in the evaluation website, five of the eight participants who took a *searching and consulting* approach were unable to find the relevant guidance in WCAG 2.0. For example, WD6 identified a potential accessibility issue relating to the audio that plays automatically upon loading the evaluation webpage but was unsure which of the four WCAG 2.0 principles contained the relevant information. The participant said:

> It's definitely not to do with things being Presentable [sic]. It's not really to do with things being Operable. Maybe to do with things being Understandable? I don't think it's to do with things being Robust. I mean, it's playing audio, which is incredibly annoying, but that's not going to impact the robustness of the interpretation of the content, I don't think. Which means it might be to do with things being understandable, possibly … There's nothing I can see in the guidelines that would indicate that playing audio might be confusing to users or disorientating.

In fact, audio that plays automatically is addressed by WCAG 2.0 success criteria 1.4.2 Audio Control, which falls under Principle 1 (Perceivable).

WCAG 2.0's role as a technical standard necessitates a certain degree of accuracy and precision; however, its domain-specific language and terminology made it difficult for some participants to search the resource for accessibility information. Having identified potential accessibility issues in the evaluation website, six of the eight participants who took a *searching and consulting* approach struggled to compare their description of the issue to the terminology used in the resource. For example, WD6 identified a potential accessibility issue relating to the videos on the webpage but was unsure which of the WCAG 2.0 guidelines were applicable. The participant said:

> It sounds like it might be 'time-based media' but I don't really know what the phrase 'time-based media' means … Media is audio or video, or a combination of audio and video. Time-based audio or video means? … I would think that all media is time-based, except for photos? So time-based audio or video is just audio or video? … The title of that guideline is 'time-based media', but everything inside that guideline is to do with audio or video … I guess I don't understand why that guideline isn't called 'audio or video'.

Despite WCAG 2.0 being a comprehensive source of accessibility information, the large volume of information it contains hindered some participants. Having identified potential accessibility issues in the evaluation website, six of the eight participants who took a *searching and consulting* approach struggled to locate specific accessibility information in WCAG 2.0, which they described as daunting and overwhelming. In lieu of a search function in WCAG 2.0 (which is recommended in WCAG 2.0 Technique G161: Providing a search function to help users find content), three participants (WD4, WD30 and WD31) each instead used their browser's text search feature to interrogate the resource. Rather than locating the relevant information, however, this action took the participants to the Glossary, where they gained detailed definitions of terms but no guidance on how to address the accessibility issues. WD30 also made several unsuccessful attempts to locate information using the WCAG 2.0 Table of Contents. Despite acknowledging that the information that WCAG 2.0 contains is "all useful", WD30 concluded that the resource is: "not very user-friendly considering it should be the grassroots source of information."

8.3.2.2. Searching and consulting using WebAIR

The four participants who took a *searching and consulting* approach to the website accessibility evaluation task using WebAIR identified between 21 and 25 accessibility problems (mean: 23.3 problems; SD: 1.71).

The effectiveness of WebAIR's content-oriented structure and organisation depended on the granularity of information sought. Participants looking for accessibility information about a particular type of content (e.g. forms, images, text etc.) were more successful than those searching for specific accessibility issues. Having identified broad areas of concern in the evaluation website, each of the four participants who took a *searching and consulting* approach located the relevant section of WebAIR and worked through the corresponding questions. Both WD3 and WD10 commented that WebAIR acts as a good prompt in such circumstances, reminding them of additional issues to consider. Participants who identified specific accessibility issues in the evaluation website, however, struggled to locate the relevant guidance in the resource. For example, WD3 identified a potential accessibility issue in the evaluation website relating to the use of HTML **`legend`** elements but was unable to find any information about it in WebAIR. Though WebAIR provides such information, it is not explicitly referred to by its corresponding top-level question: *Are there groups of inputs (e.g. checkboxes/radio buttons) that are grouped visually on the page?*

Though WebAIR's development-oriented language and terminology may have facilitated participants' assessment of the relevance and applicability of the information, it does not always speak the language of web developers. Having identified potential accessibility issues in the evaluation website, each of the four participants who took a *searching and consulting* approach managed to match their understanding of the issues to the descriptions used in WebAIR. For one participant, however, the language and terminology used in WebAIR was not technical enough. WD3 identified a potential accessibility issue relating to the leading, or line-spacing, of text on the evaluation website. Being aware of the CSS property that controls the amount of space between lines of text, the participant searched WebAIR for the term "line-height". After struggling to find the relevant information, the participant asked: "I am using the right terminology, aren't I?" before abandoning the search and moving onto another issue. Though WebAIR provides information about line-spacing, it is summarised by the top-level question: *Have you provided sufficient letter-spacing, line-spacing and paragraph-spacing?*, which refers to line-spacing, and not line-height.

The reduced volume of information contained in WebAIR facilitated participants' search for, but not necessarily their comprehension of, accessibility issues. Although each of the four participants who took a *searching and consulting* approach were able to efficiently locate information in WebAIR, two of them found the clarity and detail of the information insufficient. For example, WD29 identified a potential accessibility issue relating to user-defined foreground and background colours on the evaluation website. Despite easily locating information about this issue under the top-level question: *Can the foreground and background colours of the text be changed by the user?*, the participant did not find any practical guidance on how to adjust colours in specific browsers. Similarly, having easily located information about link highlighting under the top-level question: *Does each link highlight when it has focus?*, WD3 did not find enough information to adequately address the issue and instead turned to Google for the answer.

8.3.2.3. Following and applying using WCAG 2.0

The eight participants who took a *following and applying* approach to the website accessibility evaluation task using WCAG 2.0 identified between 10 and 15 accessibility problems (mean: 12.4 problems; SD: 2.07).

Though WCAG 2.0's three levels of conformance (Level A, Level AA, and Level AAA) provided some degree of prioritisation, the resource offered few clear calls to action due to the location of practical techniques deep within its structure and organisation. Having identified potential accessibility issues to investigate from WCAG 2.0, six of the eight participants who took a *following and applying* approach struggled to apply the guidance to the evaluation website. For example, while browsing the main index page of WCAG 2.0, WD8 repeatedly stated: "There's nothing really actionable for me to work my way through." The participant explained: "These are really quite high-level things that hold a lot of meaning with what they are saying but I don't really know how to test for that or check that that's right." Having explored further, the participant concluded: "I'm sort of just jumping around, there's not really any clear direction for me other than to work my way from top to bottom and then get frustrated with each one."

For WCAG 2.0 to apply to the broadest range of user groups, devices and technologies, some degree of generalisation is inevitable. Frequent ambiguities in WCAG 2.0's language and terminology, however, made it difficult for some participants to interpret and apply the information it contains. Four of the eight participants who took a *following and applying* approach struggled to compare WCAG 2.0's titles, headings and descriptions to content elements in the evaluation website. For example, WD15 misinterpreted WCAG 2.0 Guideline 3.1 (Readable) by assuming it referred to the presentation and not the comprehensibility of text content. Similarly, both WD2 and WD8 were unsure whether WCAG 2.0 Guideline 1.2 (Time-based Media) would apply to video content. Both WD8 and WD24 commented on the ambiguity of WCAG 2.0's three levels of conformance (Level A, AA, and AAA). WD8 was unsure whether the three levels prioritised the information in ascending or descending order. Having identified a Level AAA success criterion, WD8 said: "I don't know what that means … I've seen a couple that are just AA so maybe that's how much I need to care about it on my project?"

The large volume of information contained in WCAG 2.0 facilitates learning about web accessibility but impedes using the resource as a prompt or checklist. Four of the eight participants who took a *following and applying* approach struggled to efficiently process the information contained in WCAG 2.0 within the time available. Participants were particularly frustrated by the amount of time it took to navigate the resource, given the limited amount of time they had in this task (and would have in a real-life situation). For example, WD13 said:

> If you look at this [the WCAG 2.0 homepage] you think it looks quite simple... section 1, the perceivable stuff, then we scroll down and go through that to section 2, but once you actually go into it, it's a lot to interpret. And it's a lot to interpret when it's just a part of your job … You click on the links and look at techniques and what have you got here? Another 200 links! To follow and read! How do you get to claim a full understanding of such an in-depth set of documentation?

WD8 also found reading WCAG 2.0 very time-consuming: "Every time I click somewhere, there is more and more to read … I'm spending a lot of my time reading the resource rather than actually looking at my code or in any way relating it to my project."

8.3.2.4. Following and applying using WebAIR

The thirteen participants who took a *following and applying* approach to the website accessibility evaluation task using WebAIR identified between 8 and 30 accessibility problems (mean: 16.0 problems; SD: 6.09).

WebAIR's content-oriented structure and organisation and question-based phrasing provided an obvious order or sequence for participants to follow with clear calls to action. Having identified potential accessibility issues to investigate from WebAIR, each of the 13 participants who took a *following and applying* approach were able to apply the guidance to the evaluation website. The repetition of certain questions across the different sections of WebAIR, however, resulted in ten participants unnecessarily duplicating their efforts. For example, having worked through the *Links* section of WebAIR, which includes several questions relating to text alternatives of image links, four participants (WD9, WD20, WD23 and WD25) were confused to find a similar set of questions in the *Images* section. WD20 said: "Haven't we been through images? Am I doing this wrong? I thought we'd done images already? Oh, we did images within links? Okay. I've accidentally answered a lot of these questions already." Four participants (WD7, WD14, WD20, and WD23) were also confused by the repetition of questions relating to validation errors across several sections of WebAIR (e.g. *Are the tables free of any validation errors?*, *Are the images free of any validation errors?* etc.).

WebAIR's development-oriented language and terminology made it easier for participants to interpret and apply the accessibility information. Each of the thirteen participants who took a *following and applying* approach managed to compare WebAIR's content categories, questions and descriptions to content elements in the evaluation website. The choice of terminology in some sections of WebAIR, however, caused confusion for eight participants. For example, six participants were unsure whether questions relating to "validation errors" referred to the syntactical correctness of code or to the appropriateness of submitted form data. Two participants highlighted the ambiguity of certain questions, such as: *Is the webpage text appropriate to a lower secondary school education level?*, *Is the text of a sufficient size and contrast?* and *Does the caption or summary suitably describe the purpose or contents of the table?* Other terms and phrases used in WebAIR that some participants found confusing were "variation in presentation of text" (which refers to formatting options, such as bold, italics, underline etc.) and "used consistently throughout the website" (which refers to content that is repeated across different webpages, such as navigation menus and form labels).

The reduced volume of information contained in WebAIR supported using the resource as a prompt or checklist but offered no prioritisation. Each of the thirteen participants who took a *following and applying* approach managed to process the accessibility information contained in WebAIR and apply it to the evaluation website. Without information on how to prioritise the information that WebAIR contains, however, three participants felt obliged to work through every question and sub-question in each section, irrespective of whether it applied to the evaluation website. Given the limited amount of time they had in this task (and would have in a real-life situation), some participants would have preferred more guidance on how to approach the resource. For example, WD1 said:

> If you've only got half an hour or half a day, there's no real guidance on which bits you'd want to focus on. Because it's broken up into sections you feel like you should do the whole section, like the whole bit of *Text*, whereas maybe only half of *Text* is the most important bit and then you should be moving onto, say, the *Images* section. So, something around that would have been more helpful to get the value out of it in a limited time.

8.4. Discussion

Both the quantitative and qualitative findings of this investigation indicate that the design decisions made in relation to the language, organisation and volume of WebAIR were effective in supporting web developers to undertake website accessibility evaluations.

Participants using WebAIR to evaluate the accessibility of a website identified a significantly greater number of accessibility problems than those using WCAG 2.0. They also rated WebAIR significantly more highly than WCAG 2.0 across a number of characteristics, including ease of use, navigability, understandability, clarity of organisation, and—surprisingly, given the comprehensiveness of WCAG 2.0—completeness. These results provide objective and subjective assessments of the effectiveness of the two resources. They are supported by the qualitative observations and participant commentaries, which provide a broader perspective of how the two accessibility information resources support web developers to undertake web accessibility evaluations.

Participants took different approaches to the website accessibility evaluation task. Some began from the evaluation website, inspecting both the content and the underlying code for potential problems, before actively *searching and consulting* the resource. Others began from the resource, *following and applying* the accessibility information and guidance it contains to the evaluation website. The reason for this difference in approach is unclear. It may reflect participants' existing working practices and their existing knowledge of and familiarity with web accessibility, but it could also be due to how the two resources present information. For instance, although WebAIR is not explicitly presented as a checklist for web developers to follow, it nevertheless provides an ordered sequence of items that they may choose to address in such a way. This would account for the majority of WebAIR users following the accessibility information and guidance contained in the resource and applying it to the evaluation website.

Depending on participants' approach to the website accessibility evaluation task, both WCAG 2.0 and WebAIR have strengths and weaknesses. For web developers taking a *searching and consulting* approach to website accessibility evaluation, WCAG 2.0 presents a comprehensive collection of accessibility recommendations and practical techniques. Access to this information, however, may be impeded by a confusing structure and organisation, unfamiliar domain-specific language and terminology, and insufficient means of searching and navigating the resource. A consequence of this is that web developers may shun the resource in favour of other, potentially less reliable, sources of information or rely instead upon gut instinct and guesswork.

For web developers taking a *searching and consulting* approach to website accessibility evaluation, WebAIR's content-oriented structure and organisation and development-centred language and terminology present a familiar (but perhaps not sufficiently technical), understandable and manageable body of information. Although WebAIR presents fewer barriers to web developers, the reduced volume of information it contains does not always provide sufficient detail or clarity for those looking to learn about a particular accessibility issue or to bolster their existing knowledge. A consequence of this is that web developers may fail to grasp the motivation and reasoning behind the recommendations made in WebAIR, resulting in them applying the guidance arbitrarily.

For web developers taking a *following and applying* approach to website accessibility evaluation, their efforts are assisted by the information in WCAG 2.0 being prioritised according to three levels of conformance (Level A, AA, and AAA). While such prioritisation supports time-limited web developers to address the most impactful accessibility concerns, any advantages it provides are largely negated by practical guidance on how to actually address the concerns being buried within WCAG 2.0's deeply hierarchical structure and organisation. Without any clear or immediate calls to action, web developers tend to browse the resource haphazardly, seizing upon anything familiar or recognisable, and observing only the guidance that can be easily applied to their websites.

For web developers taking a *following and applying* approach to website accessibility evaluation, WebAIR's question-based phrasing and content-oriented organisation and structure offers many calls to action. While web developers' ability to apply the information that WebAIR contains to their websites may be facilitated by its action-oriented structure, their effectiveness in doing so may be hindered by WebAIR's lack of prioritisation. Without any apparent order to the resource, web developers may not allocate their limited time appropriately, spending too much time on the less impactful accessibility concerns or attempting and failing to address every single question and sub-question in the resource.

The effectiveness of the two resources in supporting web developers to undertake website accessibility evaluations depends on whether they use them as reference tools for looking up and learning about specific accessibility issues or as prompts or checklists, against which they can evaluate websites. As reference tools, with WCAG 2.0, the information is all there, web developers just struggle to get to it; with WebAIR, web developers can easily get to the information, it just is not all there. As prompts or checklists, with WCAG 2.0, the information is prioritised but there are not enough calls to action; with WebAIR, there are many calls to action but no prioritisation.

WCAG 2.0's unrivalled comprehensiveness perhaps makes it a more effective resource for learning about the technical aspects of web accessibility but due to the difficulties it presents in terms of accessing and applying the information, it is a less effective resource for actually identifying, understanding and addressing accessibility issues. WebAIR does not always provide sufficient detail or clarity for web developers to adequately develop their knowledge of web accessibility but the advantages it presents in terms of understanding and applying the information makes it a more practical, pragmatic resource for undertaking website accessibility evaluations.

This study was designed to address several limitations of the previous study. For instance, the student participants' lack of motivation and incentive in the previous study was addressed by recruiting professional web developers, for whom web accessibility is something they *should* already be doing as part of their job. Their motivation was bolstered by an, admittedly modest, offering of £20 worth of Amazon vouchers in compensation for their time and effort. By conducting the experiment either in person, at a location convenient to each participant, or remotely, using video conferencing software, participants could undertake the task in a more realistic environment, as opposed to a university laboratory. Also, because the experiment could, essentially, be conducted anywhere, it allowed greater flexibility in recruiting participants.

The use of live websites offered a certain degree of authenticity in the previous study but decreased control over the stimuli. While the purpose-built website used in the current study may have appeared less authentic, it nevertheless provided a consistent, reliable and—for the sake of the experiment—inaccessible stimulus that could be easily controlled. Unlike the previous study, in which participants used both WebAIR and WCAG 2.0, participants in the current study used only one of the two resources. Although recruiting participants with no prior knowledge or experience of WCAG 2.0 would have been extremely difficult, the use of only one of the two resources in this experiment hopefully reduced the possibility of training effects. Finally, by asking participants to verbally describe their thoughts instead of writing them down, the methodological burden was reduced, allowing participants to do more in the time available.

Despite these methodological improvements, the current evaluation is not without its limitations. A benefit of requiring participants to use only one accessibility information resource was the increased time it allowed them to complete the evaluation task (35 minutes, an increase of ten minutes over the previous study). However, this amount of time was still not sufficient to undertake a comprehensive accessibility evaluation, a process that, in reality, may take web developers hours or even days. Consequently, the evaluation can only represent an approximation of a true website accessibility evaluation. Nevertheless, along with the other aspects of the study (i.e. participant briefing/debriefing, attitudinal and demographic questionnaires) each evaluation session lasted approximately 60 minutes, which I felt was the maximum amount of time professional web developers could be expected to give. Irrespective of the authenticity of the purpose-built website used in this evaluation, in asking participants to evaluate a website that was not their own, a further degree of artificiality may have been added to the task. Both of these limitations highlight the need for a longer, more realistic experience of WebAIR that would increase web developers' familiarity with the resource and allow them to incorporate it into their existing working practices.

8.5. Conclusions

The outcome of this study is more encouraging than the previous study. It provided another opportunity to evaluate WebAIR but under more ecologically valid experimental conditions. In addition to participants rating the resource significantly more highly across a range of attributes, WebAIR was significantly more effective than WCAG 2.0 in supporting professional web developers in identifying accessibility problems and solutions in a purpose-built website. These objective and subjective assessments of the two resources were complemented by the qualitative observations and participant commentaries. These provided a broader perspective of how the two resources support web developers to undertake web accessibility evaluations. They indicated that, irrespective of the strategies that web developers adopt, both WebAIR and WCAG 2.0 have their strengths and weaknesses. My decision to reduce the volume and comprehensiveness of WebAIR in order to tackle information and procedural overload may have limited the usefulness of the resource for web developers who are already fairly familiar with web accessibility. Further work is necessary to boost the amount of information in WebAIR for those using the resource as a reference tool and provide more structure for those using it as a prompt. However, for typical web developers, who might only have a passing knowledge of web accessibility, WebAIR appears to be a more useful, pragmatic resource that provides a much-needed complement to existing tools, guidelines and resources. The evaluation of WebAIR continues in the following chapter, in which the long-term effectiveness of WebAIR is evaluated under more realistic experimental conditions and its effectiveness in supporting the creation, in addition to the evaluation, of accessible websites is investigated.

Chapter 9. Evaluation of WebAIR's effectiveness in supporting accessible web development

9.1. Introduction

This chapter concludes both the evaluation stage of this investigation and the programme of research by presenting a final evaluation of WebAIR with professional web developers. The study focuses on the ease of use and effectiveness of WebAIR in supporting web developers to create accessible websites. Unlike the previous studies, its objective is not to compare the effectiveness of WebAIR to that of other resources but instead to gain a more comprehensive understanding of how professional web developers use the resource and whether and how it can be integrated into their existing working practices. While the use of a single group study design may pose a threat to internal validity, making it difficult to attribute any observed effects to the use of WebAIR, this is made up for by the advantages gained from a detailed evaluation of the resource under ecologically valid conditions.

The initial hope for this study was to longitudinally evaluate WebAIR by having professional web developers use the resource as part of their everyday working practices for a period of three months. A longer study duration would have not only allowed for a more in-depth evaluation of WebAIR, particularly in terms of how well the resource can be integrated into web developers' existing working practices, but also provided the opportunity to determine the impact of WebAIR upon web developers' knowledge and understanding of web accessibility – an aspect of the evaluation that was not possible to assess in the previous studies.

Unfortunately, despite an initial willingness of a good number of participants to take part, for a number of reasons, including organisational constraints, they were unable to commit to the study duration, and so the study was abandoned and re-designed. Instead of having participants use WebAIR during the development of their own websites, participants in the revised study would use the resource over a period of two weeks to develop an accessible webpage that conforms to a pre-prepared specification. This re-designing maintains the study's ecological validity without compromising participants' professional commitments.

To draw attention to web accessibility barriers faced by web developers, and also to highlight the difficulties of conducting research "in the wild", this chapter presents both the design details of the initial abandoned study as well as the design and outcome of the revised study.

9.2. Evaluating WebAIR "in the wild"

The purpose of the "in the wild" evaluation was to explore the long-term use of WebAIR. Its aim was to investigate the ease of use and effectiveness of WebAIR in supporting professional web developers to create accessible websites and to establish whether and how WebAIR can be integrated into their existing working practices.

The initial study design involved web developers using WebAIR as part of their day-to-day web development activities for a period of three months and keeping an online diary documenting their experiences. On each occasion that participants used WebAIR, they would use a short online survey to record their reflections and experiences. Approximately every two weeks throughout the study period, they would complete a separate online "checkpoint" survey in which they would rate WebAIR on a range of characteristics and leave any general comments. At the end of the study period, they would complete a short survey to reflect on their overall experiences of using WebAIR. Their knowledge and understanding of web accessibility would be assessed using a quiz and their mental models of web accessibility would be elicited using similar exploratory methods to those used in Chapter 5. The participants' diary entries, ratings, and final survey responses, would be collated and analysed to build a picture of the long-term effectiveness of WebAIR. Gaining participants' long-term commitment, however, would prove extremely challenging.

Participants were initially recruited from those who took part in the previous study. Fifteen professional web developers expressed an interest in taking part. To accommodate their work schedules (e.g. regarding their project commencement dates) and also to make the study easier to administer, participants were recruited on a rolling basis with staggered start dates. This would also provide the opportunity to monitor the study method and, if necessary, make any adjustments. Despite the initial interest in participating in the study, however, only three of the fifteen web developers consented to take part. The remaining web developers were reluctant to participate in the study for several reasons: some felt the three-month study duration was too much of a commitment; and some felt the offer of a £30 Amazon voucher was insufficient compensation for the anticipated time and effort involved. The most concerning reason provided for not wanting to participate, however, was that they anticipated they would struggle to incorporate web accessibility into their existing workflows.

Meanwhile, the three participants who had consented to take part in the study appeared to be struggling. Between them, they had only completed three diary entries in the first two weeks of using WebAIR, and none of them had completed the first "checkpoint" survey. After following this up with each of them, it transpired that they too had been struggling to incorporate web accessibility into their daily activities. For example, one participant admitted they had not had the opportunity to use WebAIR, as web accessibility was not a requirement on their current project. Another participant said they had been designing as opposed to developing websites and so had not yet had the opportunity to use WebAIR. The third participant explained that they had been working to a tight deadline and had therefore been unable to commit any resources to making the website accessible. With the three participants' inability to commit to regularly using WebAIR, and with the ongoing difficulties in recruiting additional participants, there was no other option but to abandon and reconsider the study. The rest of this chapter describes the design and outcome of this revised study.

9.3. Method

9.3.1. Design

To investigate the ease of use and effectiveness of WebAIR in supporting professional web developers to create accessible websites and to establish whether and how WebAIR can be integrated into their existing working practices, a revised "in the wild" evaluation was undertaken.

Instead of having participants use WebAIR during the development of their own websites, they were asked to use the resource to develop an accessible webpage that conforms to a pre-prepared specification. Participants were provided with a realistic webpage specification as well as a colour scheme, branding, and other content resources. This approach helped to maintain the study's ecological validity without compromising participants' professional commitments. Due to participants' previous reluctance to engage in a long-term evaluation of WebAIR, as well as the limited amount of time remaining to complete the study, the evaluation took place over a period of two weeks. While this reduced study duration would no longer reveal the long-term effectiveness of WebAIR and regrettably would not provide the opportunity to determine the impact of WebAIR upon web developers' knowledge and understanding of web accessibility, it would, nevertheless, still provide the opportunity to monitor the use of WebAIR in a realistic context.

Participants were given two weeks to use WebAIR in the development of an accessible homepage for a fictional gym operator, Witness the Fitness, and to keep an online diary documenting their experiences. They were encouraged to complete the task in their own time, using their own computer. On each occasion that participants used WebAIR, they were prompted to complete a short online survey to record their reflections and experiences. At the end of the study period, participants completed a final online survey reflecting on their overall experiences of using WebAIR. In addition to evaluating the accessible webpages that participants developed, their diary entries and final survey responses were collated and analysed to investigate the ease of use and effectiveness of WebAIR.

The revised method was piloted on a web developer who is a friend of the author, with the procedure refined in light of that experience.

9.3.2. Participants

Fifteen professional web developers took part in the study. Three participants were female, the rest were male. The age of participants ranged from 23 to 43 years (mean: 31.0). One participant was from Australia, the rest were from the UK. The majority of participants were recruited from those who took part in the previous study. Additional participants were recruited using opportunistic sampling from web developer mailing lists, online forums, social media, conference flyers and personal contacts in the industry. The participants had between 1 and 17 years of experience of web development (mean: 9.0 years). Three participants worked for large enterprises (250+ employees), eight worked for SMEs (< 250 employees) and four participants were self-employed freelancers. Participants were offered £30 worth of Amazon vouchers in compensation for their time and effort.

Participants were required to be familiar with developing websites but were not required to be familiar with web accessibility. Prior to the evaluation, participants reported being fairly familiar with web accessibility (mean: 4.0; SD: 0.76) and moderately familiar with WCAG (mean: 3.3; SD: 1.10) (both scales: 1 = Not at all familiar, 5 = Very familiar).

9.3.3. Materials

Given the nature of the evaluation, participants used their own computer with Internet access, so all communications about the study were via email.

Participants were provided with a basic website specification outlining what was required from the webpage (see Appendix G). The specification required participants to include a range of different content types (e.g. images, tables, forms, video etc.). Participants were also provided with all of the necessary content resources (e.g. text, images, videos, branding and a colour scheme). The images provided to participants were free of any copyright restrictions. The video provided to participants was produced by York Sport, a sports centre based at the University of York. Permission to use the video was granted from both York Sport and the production company that produced the video. Many aspects of the webpage design were left open to the web designers' interpretation and judgement but one of the key requirements in the specification was to make the webpage as accessible as possible, using WebAIR.

Participants conducted the web development task using the complete version of WebAIR. To remind them to complete regular diary entries, participants used a version of WebAIR that was configured to display a prompt, via a pop-up after one minute of use, to complete the feedback form. The prompt could be dismissed without completing the feedback form and, once dismissed, it did not reappear during that session.

At the start of the study, participants viewed a webpage containing information about the study and completed an online informed consent form (see Appendix A). During the study, participants had access to the diary entry survey that they could complete as many times as necessary (see Appendix E). At the end of the study, participants completed a survey relating to their experience of using WebAIR, which included a short demographic questionnaire (see Appendix F).

9.3.4. Procedure

Participants had the opportunity to read the study information webpage, before completing the first part of the online consent form. They then rated their familiarity both with web accessibility and with WCAG on two 5-point items (1 = Not at all familiar; 5 = Very familiar). At the end of this short introductory survey, participants could download a .zip file containing the study materials. This included the webpage specification document as well as all of the necessary content resources (e.g. text, images, videos and branding). The two-week study period began once participants had downloaded the study materials.

At the end of the study period, participants uploaded their completed webpages to Google Drive and completed the post-study survey. This was the same questionnaire as used in the previous three studies (see Section 6.3.3). Participants were also asked to reflect on how WebAIR fitted into their workflow, and to leave any further comments.

Finally, participants completed a brief demographics questionnaire and the second part of the online consent form, in which they provide their consent to submit their responses.

9.3.5. Data preparation

Investigating the accessibility of the participants' completed webpages proved unexpectedly difficult. As described in Chapter 6, the concept of conformance or priority levels was purposely avoided in the design of WebAIR. This was to promote user experience over standards conformance, in order to address the observation made in Chapter 5 that web developers' have a conformance-oriented mental model of web accessibility.

This meant, however, that there was no obvious way of determining the accessibility of the participants' completed webpages using WebAIR, other than by evaluating them against each of the 156 Yes-No questions in the resource. Under this scoring rubric, positively rated questions would reflect web accessibility successes and negatively rated questions would reflect web accessibility failures. Questions that did not apply to the participants' webpages (e.g. *Are captions provided where needed?*) would be rated 'Not Applicable' (N/A). The frequency and percentage of the 'Yes', 'No' and 'N/A' ratings could then be calculated to provide an overall set of accessibility scores for each webpage. Although I had not intended WebAIR to be used to establish a benchmark in this way, this summation nevertheless provided a means of quantitatively comparing the accessibility of the participants' websites. To provide a more indicative benchmark of web accessibility, participants' completed webpages were also audited against WCAG 2.0.

Of the 156 questions in WebAIR, 23 are neutral questions for filtering/branching purposes (e.g. *Are the images purely decorative?*). These may be rated positively *or* negatively without reflecting the accessibility of the website under evaluation. Ratings for these questions were not included in the total accessibility scores for each website. Furthermore, one of the questions (on code validation) is repeated six times across different content type categories (e.g. *Are the forms free of any validation errors?* and *Are the links free of any validation errors?* etc.). Given the difficulty in determining the number of validation errors in a particular content type, the six code validation questions were combined into one question covering the entire webpage. Therefore, excluding the 23 neutral questions and five of the six repeated questions about code validation, this leaves a total of 128 questions from which the frequency and percentage of 'Yes', 'No' and 'N/A' ratings can be derived, and the accessibility of websites can be determined.

In addition to evaluating the accessibility of their completed websites, the participants' in-study diary entries and post-study survey responses were collated. From these, both descriptive statistics and qualitative findings were derived.

9.4. Results

9.4.1. Diary analysis

Despite frequent prompts and reminders to provide regular feedback on using WebAIR, only 4 of the 15 participants generated any diary entries. Altogether, these participants generated a total of 13 diary entries, however, the majority (10) of entries were made by a single participant (WD10). In lieu of more prolific diary entries from participants, this single participant's diary entries will serve as a case study of their particular experience in using WebAIR to create an accessible website. The remaining 14 participants' experiences will be investigated only from their post-study questionnaire data.

Participant WD10 works for an accessibility company in Australia as a front-end web developer. He has 17 years' experience of working in the field of web development and reported being very familiar with both WCAG and web accessibility in general.

The participant's initial diary entry was made prior to undertaking the web development task. Having indicated that he was using WebAIR 'for general information/educational purposes', the participant explained that he was having "…a quick look to see how I want to use WebAIR while building the sample project". The participant did not look at any section of WebAIR in detail but instead attempted to gain an overview of the entire resource. He concluded: "I think I will use it for quality control checks at various points e.g. when I build a form, I can review that quickly after my first pass putting the markup together."

Several days later, and over a period of approximately 2.5 hours, participant WD10 made a further 9 diary entries documenting his use of WebAIR. In the majority (7) of these entries, the participant indicated that he was using WebAIR for the purpose he had claimed he would: 'to evaluate a particular content type'. The particular content types he explored were: forms (generating 3 diary entries), tables (1 entry), and images (3 entries). In another entry, the participant indicated that he was using WebAIR 'to look up a specific accessibility issue' (which related to grouping associated links in a list), and in another entry he indicated he was using WebAIR 'to evaluate an entire webpage/website' (which related specifically to checking the readability of websites).

In 7 of his 10 diary entries, participant WD10 indicated that he had been unable to find the information he had been looking for in WebAIR. Closer inspection of these entries revealed that despite locating the relevant guidance in WebAIR, the participant had found the clarity or level of detail insufficient to confidently address the issue. For example, in one entry, the participant stated:

> I found the advice easy enough, but not a definitive answer. In my form I have chosen to identify required fields with HTML5 only and added '(optional)' to the labels for fields which are not required. I'm not sure from WebAIR whether this is considered an acceptable practice.

Similarly, in another entry, the participant had wondered whether a particular image replacement technique was accessible or not but had struggled to find a satisfactory answer in WebAIR. In a diary entry concerning readability of text, the participant found information about the issue in WebAIR sufficiently detailed but not practical or specific enough. The participant said he,

> would like some simple advice on how to test and fix reading level. The advice is all great, but labour intensive. Checking the readability score in [Microsoft] Word required Googling to find out how to enable a setting under spelling and grammar checks. Lowering this barrier will be necessary to engage developers I feel!

Similar comments were made by participants in the evaluation presented in Chapter 8, who found that the reduced volume of information contained in WebAIR does not always provide sufficient detail or clarity for those looking to learn about a particular accessibility issue or to bolster their existing knowledge.

In participant WD10's final diary entry, he raised an unexpected issue regarding the question-based phrasing used in WebAIR. After reading only the top-level question *Are there any images on the page that flash or blink rapidly?* and the follow-up questions *Do they flash less than three times a second?* and *Do they blink for less than 5 seconds?*, the participant was unsure about the definition of the terminology used. In his penultimate diary entry, he said:

> I need more info to evaluate whether the carousel I am using is failing this criteria. It is mostly still, for 4-5 seconds, then there is a transition to the next slide. The transition is animated, I'm not sure of the exact frame rate. Does that count as a flash? Is there a threshold of safe changes or is any change in the page an issue?

Having further explored the relevant guidance in WebAIR, which directs participants to download and run the Photosensitive Epilepsy Analysis Tool (PEAT) developed by the Trace Research & Development Center[70], the participant followed up on his previous entry: "OK, the PEAT tool handled my last feedback nicely! I should have read the content better." The participant went on to explain the source of his confusion: "I guess the headings like 'Do they flash less than three times a second?' make me feel like I should self-assess. Maybe it should be more like 'Does all movement on the page pass the PEAT test?'" As participant WD10 points out, phrasing the accessibility issues covered by WebAIR as concrete, actionable questions could cause web developers to make subjective (and possibly uninformed) judgements without consulting the guidance in the associated 'More Information' pages. This behaviour is reminiscent of how

[70] https://trace.umd.edu/peat

participants in the laboratory study, described in Chapter 7, approached the accessibility evaluation task. Instead of consulting the accessibility information resources, they largely drew upon their own (limited) knowledge and judgement to assess each accessibility issue.

9.4.2. Accessible web development task

Each of the participants completed the web development task and created a functional homepage for the fictional gym operator, Witness the Fitness. Examples of two of the participants' completed webpages are shown in Figure 9.1 and Figure 9.2. Note from these two examples how, despite working from the same webpage specifications brief and using the same set of content resources, both webpage designs are substantially different. Screenshots of all participants' (very different) designs are provided in Appendix G.

Some participants followed the webpage specification document more closely than others. Of the various content and features outlined in the specification, eight participants neglected to present the provided images dynamically (e.g. via a carousel or accordion), instead presenting them statically. Three participants neglected to include a search facility (even though the specification did not require this to be functional). One participant neglected to include the provided map (indicating the location of the gym). One participant neglected to include either the provided list of facilities, the video of the gym facilities, or the table of membership packages. It is unclear whether these omissions were due to the perceived amount of work involved in creating such features or the perceived difficulty in making such features accessible. Participants implemented all other content and features correctly.

Figure 9.1: Participant's completed webpage created as part of the accessible web development task (1 of 2)

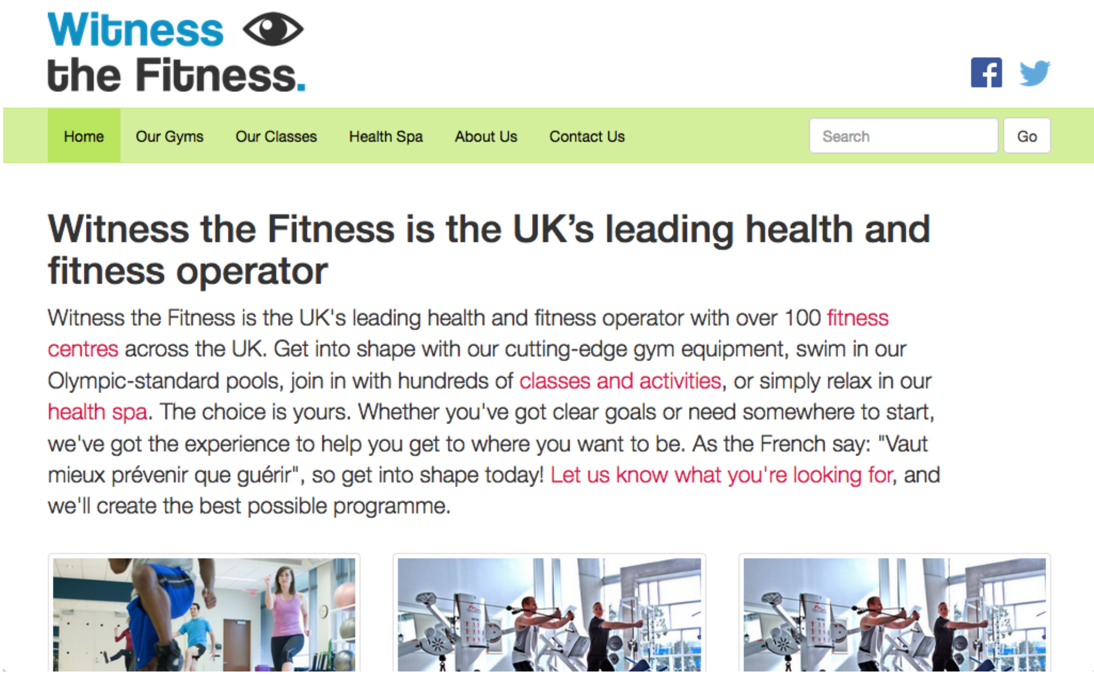


Figure 9.2 Participant's completed webpage created as part of the accessible web development task (2 of 2)

None of the participants' webpages were completely free of accessibility failures. Table 9.1 presents the aggregated scores and percentages for positively rated, negatively rated and not applicable rated questions. On average, 64 (50%) questions in WebAIR did not apply to the participants' webpages (range: 58 to 74 out of 128). This was largely due to the particular makeup of content and features in the website specification as well as the design decisions made by participants (e.g. failure to include a table in the webpage would render the questions relating to tables not applicable). On average, participants' webpages were positively rated on 45 (35%) questions in WebAIR (indicating web accessibility successes) (range: 24 to 58 out of 128) and negatively rated on 19 questions (15%) (indicating web accessibility failures) (range: 11 to 43 out of 128). With the exception of accessibility issues relating to video accessibility, which, due to the work involved, was perhaps beyond the scope of the task, all of the participants' completed webpages conformed to, or were very close to conforming to, WCAG 2.0 Level A. Negatively rated questions (indicating web accessibility failures) were fairly evenly spread across the different content type categories of WebAIR.

Table 9.1: Descriptive statistics of positively rated/negatively rated/not applicable rated questions

Question Rating	***Mean (%)***	***Standard Deviation***	***Minimum (%)***	***Maximum (%)***
YES (Accessibility Successes)	45 (35%)	8.63	24 (19%)	58 (45%)
NO (Accessibility Failures)	19 (15%)	7.45	11 (9%)	43 (34%)
N/A (Not Applicable)	64 (50%)	3.97	58 (45%)	74 (58%)

Table 9.2 presents the questions that were rated negatively for the majority (eight or more) of participants' webpages. Two of the 128 questions in WebAIR (the top two rows in Table 9.2) were rated negatively for all fifteen participants' webpages. Each of the participants' webpages failed these two questions for the same reason. The written copy provided to participants in the webpage specification document included a passage containing unusual jargon relating to different yoga practices. Though it could be argued that the participants were only working to the specification document, none of the participants included definitions or clarifications of the jargon or provided links to a glossary or list of terms. A further twelve questions (the remaining rows in Table 9.2) were negatively rated for the majority (eight or more) of participants' webpages. The participants' webpages failed these twelve questions for two distinct reasons: participants either had failed to implement accessible content or features (e.g. by not providing alternative text for images or a form submission confirmation) or had

implemented accessible content or features incorrectly (e.g. by providing text that is not of a sufficient size or contrast or code that has validation errors).

Table 9.2: WebAIR questions rated negatively for majority of participants' webpages

Content Category	***Question***	***Frequency of participants' webpages rated negatively***
Text	Do you provide definitions for any unusual jargon or slang in the content of the webpage?	15
Text	Do you provide a link to a glossary or a list of terms (including pronunciations)?	15
Text	Is the text of a sufficient size and contrast?	14
Text	Do you provide a link or control to an alternative version of the page with text that has sufficient size and contrast?	14
Text	Do you provide controls on the webpage that allow users to change the size of the text up to 200 percent?	14
Forms	Do you provide a checkbox in addition to a submit button for users to confirm their responses before submitting them?	13
Forms	Do you provide feedback when a form has been submitted successfully?	13
Audio & Video	Do you provide an audio alternative (e.g. a second audio track) which describes the video contents?	12
Audio & Video	Does all audio and video content have a text alternative that describes the actual content (e.g. a transcript)?	11
Forms/Links/Tables/ Images/Text/ Audio & Video	Are the [Forms/Links/Tables/Images/Text/Audio & Video] free of any validation errors?	11
Navigation	Do you provide context-sensitive help (including explanatory text)?	9
Navigation	Do you provide information about the user's location within the website (e.g. breadcrumb trails, site map, navigation menu etc.)?	8
Images	Do all images have a text alternative (including where image replacement has been used)?	8
Images	Do you provide an additional long description (either in the text near the non-text content or a link to another location)?	8

A more complete understanding of these results, including greater clarity about the causes of the web accessibility failures found in the evaluations, is provided by the participants' responses to the post-study survey.

9.4.3. Post-study survey

9.4.3.1. Amount of time

Given the considerable time and budgetary constraints within which web developers must work, it is important that using WebAIR does not take too much time. Participants' estimates of how long it took them to develop an accessible webpage using WebAIR varied substantially, from as little as two hours to as much as three days. The mean amount of time participants estimated spending on the task was 11.8 hours (SD: 17.57) but excluding two outliers (who estimated they spent 24 hours and 72 hours, respectively) reduced this to 6.3 hours (SD: 3.56). Given the relative simplicity of the task, the latter seems like a more reasonable amount of time.

Participants rated on a scale of 1 to 5 (where 1 = Not at all, 5 = Very much) the extent to which making the webpage accessible using WebAIR had added to the time it would normally take. A one-sample t-test showed the mean rating of 2.5 (SD: 0.83) to be significantly below the midpoint of the scale (3.0), $t(14) = -2.17$, $p = .048$, $d = -0.56$, indicating that participants felt using WebAIR had added very little to development time. One participant described how "the time spent using WebAIR was made up for in terms of getting answers quickly and without having to search the web or trawl through W3C documentation" $_{[WD1]}$. Table 9.3 describes the development activities that participants felt consumed the most amount of time during the web development task.

Table 9.3: Perceived time-consuming activities during the web development task

Activity	*Frequency of mentions*
Making forms accessible	6 participants
Making images accessible (in particular, providing text alternatives)	6
Making videos accessible (in particular, providing text alternatives)	3
Making tables accessible (in particular, providing summaries of content)	3
Making links accessible (in particular, providing accessible link text and title attributes)	2
Making text content accessible (in particular, marking up the text and checking the reading level)	2
Making animated content (e.g. image carousels and accordions) accessible	2
Checking colour contrast	2
Ensuring all content is resizable	2
Making JavaScript-based interactive components accessible	1

Participants found it time-consuming making forms and other interactive components accessible. Some felt that any interactivity “requires a lot of testing” [WD14] and that, in particular, forms “need to be configured more specifically (error messages etc.)” [WD13]. The difficulty in creating accessible forms was raised in one of WD10’s diary entries, in which he criticised the lack of definitive answers in WebAIR regarding the correct use of required fields. These attitudes also correspond to the outcome of the web development task, in which the majority of web developers failed to implement any confirmation that a form had been submitted successfully. Though some participants provided custom error messages for the forms in their webpages, the majority relied solely upon the HTML5 **`required`** attribute, which automatically generates error messages specifying that input fields must be filled out before submitting the form. Although the **`required`** attribute is accessible, not all web browsers currently support this feature of HTML5.

According to one participant, animated content, such as image carousels or accordions, also requires a lot of work because “many of the plugins don't include the required accessible functionality so require some bespoke changes” [WD6]. The difficulty in making animated content accessible was raised in two of WD10’s diary entries, in which he struggled to determine what constitutes flashing or blinking content. Given that the majority of participants chose to present static, rather than dynamic, images in their websites, it suggests that making animated content accessible was perceived by these web developers as requiring too much effort.

Many of the time-consuming activities that participants mentioned relate not to coding the webpage but to providing accessible content (e.g. text alternatives, descriptive link text, and table summaries). One participant described how providing alternative text was “time-consuming” and that it “often feels verbose adding the alt and title attributes to elements. Particularly because these are not usually requested from the client so (and as in this case) are not provided and have to be written by the developer” [WD1]. These attitudes correspond to the outcome of the web development task in which many of the negatively rated questions were due to web developers failing to implement accessible content and features (as opposed to implementing them incorrectly).

As well as taking extra time, some participants described how making the webpage accessible had forced them to compromise their design. According to one participant:

"I had to make some forms pretty ugly to get them to be accessible (the search form was originally simplistic, but I realised I could not do it in the way I wanted)" [WD5]. Another participant said: "The consideration given to modern trends in web development slowed me trying to come up with a best solution for both accessibility and design. Design won in the end" [WD8].

9.4.3.2. Characteristics of WebAIR

Figure 9.3, below, shows the mean ratings of various characteristics of WebAIR.

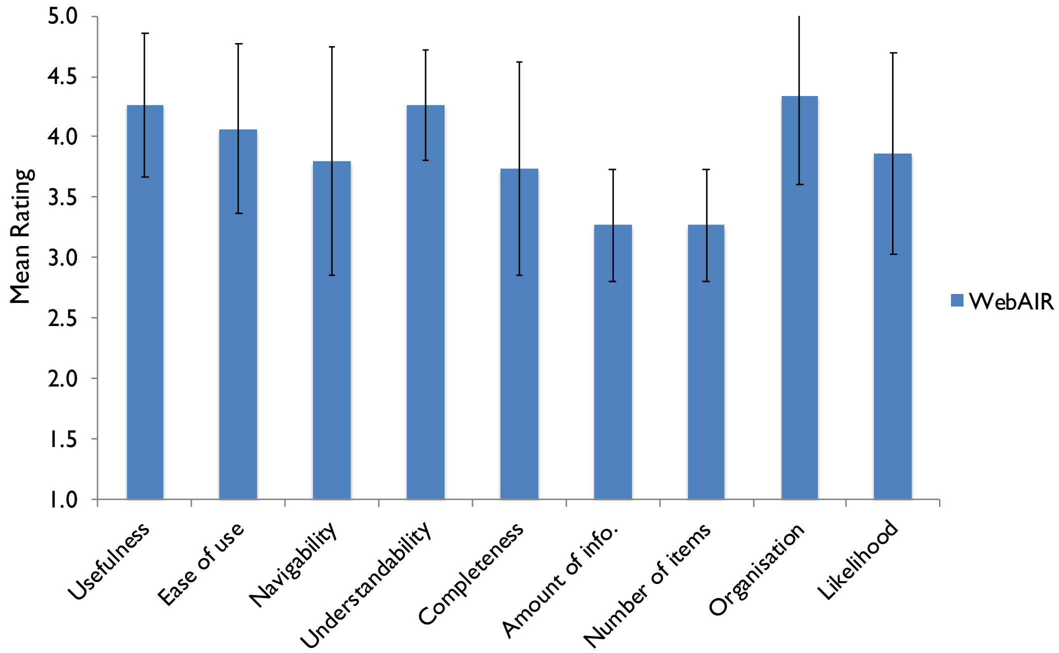


Figure 9.3: Mean ratings of WebAIR on a range of characteristics

Error bars: standard deviation

A series of one sample t-tests investigated whether the mean ratings deviated significantly from the neutral midpoint of the scale (3.0). Descriptive statistics and significance tests are presented in Table 9.4. Participants rated WebAIR fairly highly across the range of characteristics, but in particular for *usefulness*, *ease of use*, *understandability*, and *organisation*, whose mean ratings were all higher than 4 (out of 5) and significantly above the midpoint of the scale. Mean ratings of WebAIR for *navigability*, *completeness* and *likelihood of using* were slightly lower, between 3.5 and 4, but still significantly above the midpoint of the scale. Mean ratings of *amount of information* and *number of items to test* were lower, both at 3.3, but still significantly above the mid-point of the scale, suggesting that participants found the volume of WebAIR to be still bordering on too large. Effect sizes for these analyses were found to exceed Cohen's (1988) convention for medium (d = 0.50) to large effects (d = 0.80).

Table 9.4: Descriptive statistics and significance tests for deviation of mean ratings of WebAIR characteristics from the midpoint of the scale

Characteristic	*Mean*	*SD*	*t (all tests df = 14)*	*Sig.*	*d*
Usefulness	4.3	0.59	8.26	p < .001	2.13
Ease of use	4.1	0.70	5.87	p < .001	1.52
Navigability	3.8	0.94	3.29	p = .005	0.85
Understandability	4.3	0.46	10.72	p < .001	2.77
Completeness	3.7	0.88	3.21	p = .006	0.83
Amount of information	3.3	0.46	2.26	p = .041	0.58
Number of items to test	3.3	0.46	2.26	p = .041	0.58
Organisation	4.3	0.72	7.14	p < .001	1.84
Likelihood of using	3.9	0.83	4.03	p = .001	1.04

Participants rated on a scale of 1 to 5 (where 1 = Not at all confident, 5 = Very confident) how confident they were about the accessibility of the webpage they had created using WebAIR. A one-sample t-test showed the mean rating of 4.1 (SD: 0.46) to be significantly above the neutral midpoint of the scale (3.0), $t(14) = 9.03$, $p < .001$, d = 2.33, suggesting that participants felt very confident. One participant said of his experience of using WebAIR that it "felt like it followed the 80:20 rule nicely, in that the result isn't perfect but I was able to strike a good balance between effort and value" [WD1].

Participants also rated on a scale of 1 to 5 (where 1 = Not at all, 5 = Very much) the extent to which they felt their understanding of web accessibility had been increased by WebAIR. A one-sample t-test showed the mean rating of 3.5 (SD: 1.06) was *not* significantly above the neutral midpoint of the scale (3.0), $t(14) = 1.95$, $p = .072$, d = 0.50, suggesting that participants did not feel their understanding of web accessibility had been substantially increased. Despite this, one participant said, "I will have to shake my 'I already know a lot about web accessibility' mindset, but I learned quite a few things from WebAIR and it was a useful tool" [WD10].

Participants provided feedback on the aspects of WebAIR that they felt had either helped or hindered them during the web development task (see Table 9.5 and Table 9.6, respectively).

Table 9.5: Perceived helpful aspects of WebAIR

Aspect	*Frequency of mentions*
The content-oriented organisation	3 participants
The clearly written and understandable language	2
The short, well-structured Table of Contents	2
The relatively brief amount of content provided in the 'More Information' pages	2
The inclusion of code examples	2
The use of headings and boxed-off sections to break up the content	2
The question-based phrasing of accessibility issues	1
Links to further reading and other external resources	1
The presentation of content on a single page	1
The minimal colour palette	1

Participants found WebAIR's content-oriented organisation made the resource easy to navigate and to use as a checklist. Organising the content in this way made WebAIR "read like a process tree: 'do this, then do this if you have it'" [WD13] and "enabled me to check one section at a time" [WD7] which "made sense for how I think of the page" [WD1]. One participant suggested that WebAIR could provide a better sense of progression by including a 'checkbox' feature that would allow web developers to 'tick off' the questions they have answered. Another participant said they found WebAIR "much more useable than I thought I would. At first it seemed like a lot of content but referring to the relevant section when it came up was actually pretty easy" [WD5].

WebAIR's "simple language" [WD13] and terminology appeared to help participants during the web development task. One participant felt that "All the information was clearly written and understandable, and that's from someone who is a junior developer and has no previous experience in web accessibility" [WD15]. Another participant considered the accessibility issues to be "well summarised" by the top-level questions in WebAIR, which they described as "very short but usually adequate without needing to click on 'More information'" [WD14].

WebAIR's reduced volume of information also appeared to help participants during the web development task. One participant described the information in WebAIR as "detailed enough to be useful, whilst still being concise enough to quickly digest" [WD12]. Another participant said they "love the fact the main detail is all on one page" adding that it “makes it really quick to run down the list and then just click through when you need some help or more information” [WD6].

Table 9.6: Perceived hindering aspects of WebAIR

Aspect	***Frequency of mentions***
The presentation of content on separate ‘More Information’ pages	6 participants
The lack of code examples or links to further reading on some ‘More Information’ pages	4
The lack of prioritisation or order to the accessibility issues presented	4
The relatively large amount of information presented on the main page	3
The insufficient level of detail about certain accessibility issues	3
The presentation of long lines of text with insufficient line spacing	1
The complexity of the recommended further reading and other external resources	1
The lack of a visual indicator of progress through the resource	1
The lack of examples as to what the ‘accessible experience’ should be	1
The lack of syntax highlighting in the code examples	1

Many of the hindrances cited by participants related not to WebAIR’s content but to its structure and design. For example, six participants felt the inclusion of ‘More Information’ on separate pages added an unnecessary navigational burden to using WebAIR. One participant described how they had been “somewhat slowed by not having access to all the information in a single-page version” [WD2]. Another said:

> It was hard to tell from the questions which ones needed to be visited and which didn’t … but I found myself always resorting to the More Information because sometimes it was important and much more informative than the bulleted point [WD15].

One participant suggested having the ‘More Information’ pages in pop-up ‘light boxes’ or expandable concertina sections, which “might save a bit of time and prevent getting lost when you're working through a series of accessibility criteria” [WD7].

The order of the content type categories in WebAIR also hindered participants during the web development task. One participant said they “found it difficult to decide which sections to focus on first but went with Text as it makes up the majority of the site” [WD1]. One participant, WD5, wondered whether the order of WebAIR could be improved by having it roughly reflect the order that elements appear in web documents (such as by having the section on navigation come first). Several participants found it difficult to ascertain the relative importance of each category. Two participants, WD1 and WD3, suggested defining a common set of questions that would apply to the majority of webpages, though one acknowledged that the variety of webpage content would make this difficult to determine. Another participant, WD16, felt it would be beneficial to allow web developers to filter the questions in WebAIR according to the types of content in a particular website.

Another aspect of WebAIR's design that participant felt hindered them during the web development task was the visual presentation of text. One participant described how "the content is fine" but that "the information could be better presented" [WD8]. Participants complained that long lines of text with insufficient line spacing made WebAIR difficult to read and pick out information from. Suggestions for improvements included increasing the line spacing between bullet points, removing underlines from links (particularly those in the Table of Contents), colour-coding the "if yes" and "if no" selectors, highlighting syntax in the code examples, and emboldening keywords. One participant said: "Design matters: you could make WebAIR the first well-designed and accessible accessibility tool!" [WD3].

9.4.3.3. Information in WebAIR

Participants provided feedback on information that they felt was either missing from WebAIR or was in WebAIR but which they considered unnecessary or irrelevant (see Table 9.7 and Table 9.8 respectively).

Table 9.7: Information perceived missing from WebAIR

Information	***Frequency of mentions***
No information currently missing from WebAIR	5 participants
Code examples and links to further reading	4
Up-to-date guidance on HTML5, CSS3 and, in particular, WAI-ARIA	4
How to create accessible page layouts, particularly regarding responsive and resizable content	2
How to make accessible dynamic content	1
How to use accessible design patterns and widgets (e.g. tabs, carousels, accordions etc.)	1

One of the design decisions made in the creation of WebAIR was to include code examples and links to further reading only where necessary to supplement the existing information. Participants found these features very helpful during the web development task, with four participants indicating that they would prefer code examples and links to further reading to be included on every 'More Information' page. This was also suggested by novice web developer participants in the initial evaluation of WebAIR, described in Chapter 6. While more prolific links to further reading may be possible in future revision of WebAIR, care must be taken to ensure that they link to resources that are of an appropriate level of complexity. One participant complained: "Occasionally, there were pages that didn't have examples, so I clicked on the external links but gave up on them because they were too long/complex" $_{[WD3]}$. It is also important to ensure that links to further reading supplement, but do not replace, the information contained in WebAIR. Two participants, WD6 and WD13, described how they had been forced to follow links to further reading because they had been unable to find a satisfactory answer from WebAIR.

An apparent concern of many participants was the accuracy of the information contained in WebAIR and whether or not it is up-to-date. Several participants acknowledged the rapidly-changing nature of web development, with four participants highlighting the need for WebAIR to provide up-to-date guidance on the latest standards and specifications, such as HTML5, CSS3 and WAI-ARIA. Though WebAIR has included such information since its first major revision and, as far as known, is correct and up-to-date, web developers evidently need more reassurance that this is the case. To this end, one participant suggested providing details of when WebAIR was last updated "so I can confirm that I have an up-to-date answer and don't need to necessarily double-check things" $_{[WD1]}$.

Table 9.8: Information in WebAIR perceived unnecessary or irrelevant

Information	***Frequency of mentions***
No unnecessary of irrelevant information currently in WebAIR	11 participants
How to increase the size of text up to 200 per cent	1
How to make accessible audio and video content	1
Guidance that requires human judgement	1
Information that is duplicated across multiple sections (e.g. regarding validation)	1
Guidance that is difficult to implement	1

Participants' suggestions for unnecessary or irrelevant information in WebAIR reflect an apparent lack of responsibility for some aspects of web accessibility. For example, one participant, WD5, suggested removing the section on how to make accessible audio and video content because they felt techniques such as providing captioning, audio description, and sign language translation, should be provided by third-party services. Similarly, WebAIR's recommendation to allow users to change the size of the text up to 200 percent is, according to one participant, an "outdated practice" as "browsers allow for this themselves" [WD12]. These attitudes correspond to the outcome of the web development task in which many of negatively rated questions were due to web developers failing to implement accessible content and features (as opposed to implementing them incorrectly).

Other suggestions of unnecessary or irrelevant information in WebAIR related to the resource's structure and organisation. Some participants were confused by the repetition of certain questions across the different sections of WebAIR (e.g. the question on code validation, which is repeated for six different content types, or questions on text alternatives for images, which are repeated in both the Images and Image Links sections). One participant said: "I think it's important to cut out anything that seems repetitive or differentiate it a bit more clearly" [WD15]. Another participant felt that information should not be removed from WebAIR but instead reordered to avoid unnecessary effort. Citing the example of text alternatives for images, the participant highlighted how several questions regarding the accuracy and readability of text alternatives come before the question that determines whether or not a text alternative is actually necessary.

9.4.3.4. Circumstances of use

Participants provided feedback on whether or not they would use WebAIR and under what circumstances. While all of the participants expressed an interest in using WebAIR, the circumstances in which they felt they would use the resource varied.

Three participants indicated they would use WebAIR from the very beginning of development. One explained:

> I work as a freelance web developer and like the idea of making more of an effort for my work to be as accessible as possible, especially as it increasingly becomes the norm for attention to be paid to those kind of details in development and UX [WD15].

Another participant said they would use WebAIR to educate and inform clients upfront: “I’d copy the guidance about writing copy and send it to a client as early as possible, so that they can understand it before writing their copy” [WD1]. Similarly, another participant felt that WebAIR would be “a useful resource for more junior team members to understand some of the principles and why things are done as they are” [WD6].

Four participants indicated they would use WebAIR during development, particularly as a reference tool to look up easily forgotten information. One explained: “It's easy to forget some of this stuff, especially if you're not doing it all the time, so having a clear list with useful hints and tips is a great idea!” [WD7]. Another participant, who is primarily a back-end web developer, said: “I’m going to use WebAIR to review code sent through from front-end developers to me before implementation into the solution, this way I can help prevent markup reaching production that may be less than ideal for a user” [WD12].

The most common response, mentioned by five participants, was that they would use WebAIR as a checklist towards the end of development to ensure they have “covered all bases” [WD7] and “considered everything” [WD14]. One explained:

> I’d use WebAIR towards the end of a project during the cleanup stage. Just because I'd find it very hard to justify the value of it early on when there's a rush for everything else, and I know that my fixes would have the best chance of not being broken by other developers if they went in at the end [WD1].

Several participants expressed caveats regarding the use of WebAIR. For example, one participant said they would only use WebAIR “where sensible in the situation. For example, in a content management system, it might not be possible to do all it asks” [WD13]. Another participant, WD16, felt they might use WebAIR when starting a new website but would be reluctant to use it for retrospectively addressing accessibility issues in existing websites. One participant, WD5, said they would be keen to use WebAIR for every development project, providing the budget was available. Similarly, another participant felt they would be more likely to use WebAIR for public sector websites, because “they tend to have the highest accessibility requirements and are willing to spend a little extra budget to improve things” [WD6].

9.5. Discussion

This chapter has presented a final "in the wild" evaluation of WebAIR with 15 professional web developers. The aim of the study was to gain a detailed understanding of how WebAIR is used by web developers and whether and how it can be integrated into their existing working practices. Despite a disappointing setback that resulted in the initial study being abandoned, the re-designed study demonstrated the ease of use and effectiveness of WebAIR in supporting web developers to create accessible websites. Though very few participants contributed to the diary study aspect of the evaluation, all of the participants were effective in using WebAIR to create a (mainly) accessible webpage and all of them provided valuable feedback on their experience in a post-study questionnaire.

As in the previous evaluations of WebAIR, described in Chapter 6, Chapter 7 and Chapter 8, participants appreciated WebAIR's development-oriented language and terminology, which they generally found clear and understandable. Despite their appreciation, however, several participants commented that WebAIR sometimes lacked the necessary authority and definitiveness to support them in confidently applying the guidance and recommendations. This issue was initially raised by one of the participants in the diary part of the evaluation, who had looked up various issues in WebAIR but found the guidance vague and indefinite. Several participants raised a similar issue in the post-study survey. They questioned the accuracy of the information contained in WebAIR and whether or not it is up-to-date. In lieu of definitive information, participants turned to the external resources provided in WebAIR as further reading, a process that they found time-consuming and, due to some resources being overly-complex, not always helpful. This apparent lack of authority and definitiveness may have also accounted for many of the web accessibility failures observed in the web development task, where web developers either neglected to implement accessible content and features or implemented them incorrectly. Although web developers may find WebAIR's language and terminology more familiar and easier to understand, they still need access to accurate, explicit and definitive guidance that will allow them to confidently make decisions about the accessibility of their webpages.

As in the previous evaluations of WebAIR, participants praised its content-oriented structure and organisation, which allowed them to efficiently locate the relevant section of the resource. The decision to deliver WebAIR as a multi-page website, however, posed an unanticipated barrier to using the resource effectively. Some participants described in the post-study survey how they had been reluctant to explore the 'More Information' pages associated with each question in WebAIR, either because the term suggests supplementary and, perhaps, nonessential information or simply because they found it a hassle to navigate back and forth from the main page and potentially lose their place in the resource. Consequently, participants tended to base their assessment on the questions alone, drawing answers from their existing knowledge of web accessibility, which was not always accurate or up-to-date. Similar behaviour was documented by one of the participants in the diary part of the evaluation, who described how the phrasing of certain questions in WebAIR made him feel like he had to determine the answers himself rather than consult the guidance. Many of the web accessibility failures observed in the web development task might have also been avoided through closer attention and adherence to the guidance presented in WebAIR's 'More Information' pages. While WebAIR's structure and organisation may make it easier for web developers to locate information, it needs to clearly distinguish between essential and supplementary information and prioritise practical information upfront.

As in the previous evaluations of WebAIR, participants generally found the reduced amount of information and number of items to test in WebAIR more manageable and easier to digest. Participants also felt using WebAIR added very little to development time. The decision to present WebAIR as a list of questions, however, appears to have biased many participants towards using the resource in a less than effective way. Some participants used WebAIR on demand as intended, to test the accessibility of particular types of content throughout the development process. Such behaviour was documented by one of the participants in the diary part of the evaluation, who referred frequently to WebAIR during the development of his webpage. Feedback from many participants, however, suggests they view the resource as a summative checklist to be completed in one go towards the end of the development process. This is evident from criticism of the repetition of content across different sections of WebAIR, which is only noticeable if the resource is used in this way. It also accounts for the persistent criticism from participants in both this and the previous evaluations that WebAIR still presents an overwhelming amount of information. As noted in the re-design of WebAIR (see Section 6.7.3), few options remain for further reducing the volume of WebAIR without compromising the quality and comprehensiveness of the guidance. Further work should perhaps therefore focus on reframing web accessibility as a creative—as opposed to remediative—process, allowing WebAIR to support web developers throughout—as opposed to at the end of—the development process.

The evaluations of WebAIR presented in this and the previous chapters demonstrates its effectiveness in supporting both the creation and evaluation of accessible websites. However, the difficulties encountered in setting up the initial "in the wild" evaluation, in which web developers would have used WebAIR as part of their everyday working practices for a period of three months, serve as a reminder of the organisational constraints that web developers face. As established in the contextual inquiry study in Chapter 4, the mental models study in Chapter 5, and much of the literature reviewed in Chapter 2, web developers often work under considerable time and budgetary pressures. Clients, management and other stakeholders have little awareness and no clear understanding of what accessible web development entails, leading to it being perceived as a confusing, overwhelming and unaffordable 'optional extra'. Of the fifteen web developers who showed interest in participating in the initial "in the wild" evaluation of WebAIR, only three consented to take part. Some were reluctant to commit to participating in a three-month study whereas others felt the study offered insufficient compensation for their time and effort. The most telling reason, however, was that some anticipated they would struggle to incorporate web accessibility into their existing workflows. Even the three participants who began the initial study struggled, for various reasons, to commit any time to making their websites accessible.

The re-designed evaluation, in which web developers used WebAIR over a period of two weeks to develop an accessible webpage confirming to a pre-prepared specification, proved more palatable to participants and provided the opportunity to evaluate the resource under ecologically valid conditions. Yet still, as noted above, many participants used WebAIR in addition to, and not as part of, their existing workflows. Also, many of the accessibility failures in the participants' completed webpages did not appear to be things that they had forgotten, but rather things that they simply had not bothered to do, such as providing alternative text or using additional tools and resources to check something (e.g. colour contrast, which requires an external tool). Certainly, this could have been because of the artificial nature of the task. Some participants might have felt that they should only do as asked; others might have felt that some of the guidance in WebAIR was above and beyond the expectations of a voluntary study. However, it may also reflect how web developers perceive their role and responsibilities. Some of the participants' responses in the post-study survey, such as suggesting that audio and video accessibility is the responsibility of third-party services (see Section 9.4.3.1), or that web browsers should handle text resizing (see Section 9.4.3.3), indicate that they feel web accessibility is not the responsibility of web developers. As highlighted in the introduction to this thesis (see Section 1.3), and discussed in the design and implementation of WebAIR (see Section 6.2), the considerable range of responsibilities makes the role of web developer difficult to precisely define and therefore difficult to adequately support. It is perhaps this role ambiguity, as well as the difficulty in providing guidance to suit everyone, that absolves some web developers, whether intentionally or not, of their responsibility to create accessible websites.

Irrespective of the usability of accessibility information resources or their potential effectiveness in supporting the creation and maintenance of accessible websites, web developers evidently still face a range of difficulties and challenges in integrating web accessibility into their existing working practices. The reluctance to use WebAIR as intended, and the persistence of existing flawed mental models, raises the possibility that an accessibility information resource *designed specifically to support web developers* is simply not a sufficient solution to this problem. As established in the introduction to this thesis (see Section 1.3), numerous stakeholders are involved in the value chain of accessible web development, many of whom may also potentially benefit from access to clear, concise and precise accessibility information. An accessibility information resource that supports a broader range of stakeholders or relates to different aspects of the development process (e.g. content creation, design, development, QA etc.) may therefore be a more effective solution. This broader approach, however, is contingent on the attitudes and actions of the individuals, teams, and organisations that commission, own and manage websites. Without an organisational commitment to encouraging web accessibility and embracing the principles of inclusive design, any initiatives to support the creation and evaluation of accessible websites are unlikely to succeed.

The biggest limitation of either iteration of this study is the quasi-experimental one-group case study design, in which only a single group of participants were involved. All fifteen web developers who participated in the study used WebAIR to create an accessible webpage and its effect was determined from diary entries completed during the study period, survey responses completed after the study period, and accessibility evaluations of their completed websites. The lack of a comparison or control group of participants who did not use WebAIR (or who used an alternative accessibility information resource, such as WCAG 2.0) poses a threat to internal validity, making it impossible to draw firm conclusions about the effectiveness of WebAIR in supporting professional web developers to create accessible websites. For instance, it is possible (though unlikely) that the study would have had the exact same outcome if participants had not used WebAIR but instead used other resources, such as WCAG 2.0, or had even conducted their own research. Prior to commencing the study, participants reported, on average, being fairly familiar with web accessibility and moderately familiar with WCAG, and could, therefore, have drawn on their existing knowledge of web accessibility to produce equally accessible webpages. Ideally, I would have implemented a comparative experimental design, allowing me to compare the effectiveness of different resources and draw firm conclusions about the effectiveness of WebAIR. Given the difficulties in recruiting participants for this study, however, as well as the limited time available in which to complete the study, such a design would have been difficult to implement. Furthermore, it is already well established that web developers struggle to create accessible websites using existing resources (see Section 2.5.3) and so any comparison or control group in this study would likely have only served to reaffirm this observation.

Any methodological weaknesses that threaten the study's internal validity are hopefully offset by its high ecological validity. Participants completed the study under realistic conditions: using their own web development environment, in their own working environment, and in their own time. The only artificial aspects of the study were the nature of the webpage that participants were required to develop and the fact that it was not part of their actual work. Whereas in the original study participants would have used WebAIR in the development of genuine client websites, participants in the re-designed study used the resource in the development of a webpage for a fictitious gym operator. Despite a small monetary incentive being provided, the absence of a genuine client could have affected participants' motivation to perform the task to the best of their ability. Participants knew they were not developing a real website and may have been disinclined to put in the level of effort that they might otherwise have done. Certainly, many of the accessibility failures observed in their completed webpages were due to a reluctance to provide additional content beyond that included in the webpage specification document (e.g. text alternatives, descriptive link text, or table summaries).

9.6. Conclusions

The outcome of this final study is bittersweet. On the one hand, participants demonstrated the ease of use and effectiveness of WebAIR and provided confirmation that the design decisions taken in the development of the resource are effective in supporting the creation of accessible websites. Regrettably, the shortened duration of the re-designed "in the wild" evaluation afforded little opportunity to determine the long-term impact of WebAIR both upon participants' working practices and their knowledge, comprehension and mental models of web accessibility. Nevertheless, participants were effective in using WebAIR to create a (mainly) accessible webpage and provided valuable and largely feedback on their experience in a post-study questionnaire.

On the other hand, some of the participants' responses suggest there is still some way to go in shifting web developers' concept of web accessibility from merely satisfying a collection of technical requirements to meeting the needs of as many users as possible. Participants tended to use WebAIR, not to test the accessibility of particular types of content during the development of their webpages, as intended, but, instead, to evaluate the accessibility of their entire webpage towards the end of the development process. Some also called for a means of prioritising WebAIR's content (be it by severity, importance, or job role, e.g. front-end development, back-end development etc.), much like the Level A, AA and AAA conformance levels of WCAG 2.0. Of particular concern was the notion that certain aspects of web accessibility are somehow not the responsibility of web developers. Whether this notion reflects the artificial nature of the (revised) experimental design or ambiguity over the exact role and responsibilities of web developers, the consequences are clear: however effective a resource such as WebAIR may be in supporting the creation and evaluation of accessible websites, without the responsibility and commitment, both of web developers and other stakeholders in the value chain of accessible web development, its impact will be severely limited.

Chapter 10. Overall Discussion and Conclusions

10.1. Summary of results

The programme of research presented in this thesis aimed to better understand the difficulties that experienced and novice web developers face in creating and maintaining accessible websites. In addition, it aimed to establish the necessary support to enable web developers to better integrate web accessibility into their existing workflows. A further aim was to investigate the effectiveness of this support in a series of evaluations conducted under laboratory and real-world conditions. To achieve these aims, this programme of research was divided into three broad stages: analysis, intervention, and evaluation.

The first stage—**analysis**—examined the current state of web accessibility and drew partly upon ethnographic methods to explore the current working practices of professional web developers.

The first study within this stage, presented in Chapter 3, was a systematic review of the web accessibility evaluation literature. It addressed the research question: *Has the accessibility of websites improved over the last two decades?* The results showed a persistent occurrence of accessibility problems that, despite numerous initiatives to support, encourage and compel web developers to fulfil their responsibility to develop accessible websites, does not appear to be improving. Between 1999 and 2014, on average, less than a third of evaluated websites conformed to WCAG 1.0 Level A, the most basic level of web accessibility. The study also highlighted a considerable degree of variability and noise in the data that precluded a more precise analysis.

With the first study having established the poor state of web accessibility, the following study, presented in Chapter 4, examined one of its potential causes. Focusing specifically on the stakeholders with the greatest responsibility for web accessibility, this described a contextual inquiry investigation with 13 professional web developers. It addressed the research question: *How do web developers' current working practices support or hinder the creation of accessible websites?* The study examined web developers in context to gain a greater understanding of their current working practices and to establish how web accessibility may be integrated into their existing workflows. The results indicated that web developers are hindered, not by limited awareness or concern, but by a lack of knowledge and practical guidance on how to make websites accessible. Existing tools, guidelines and resources are letting web developers down by not providing them with the support and information they need.

To more comprehensively understand the nature of web developers' confusion and uncertainty over web accessibility, the contextual inquiry investigation was followed by an interview study, presented in Chapter 5, with 26 professional web developers. This looked to elicit web developers' mental models of web accessibility and addressed the research question: *How do web developers' existing knowledge and understanding of web accessibility support or hinder the creation of accessible websites?* The study identified that web developers' mental models appear to incorporate some, *but not nearly enough*, knowledge and awareness of web accessibility. Their knowledge of how to create and evaluate accessible websites is predicated on a very limited understanding of disabled people, the ATs they use, and the challenges they face in using the web. Underpinning their limited procedural knowledge of web accessibility is a conceptual model that appears to prioritise technical compatibility over user needs and standards conformance over user experience. This leads to web developers perceiving web accessibility as a technical, rather than human challenge, and constrains their efforts to developing websites that can be *accessed*, but not necessarily *used*, by people with disabilities.

The second stage—**intervention**—built on the research findings of the previous stage by focusing on the development and initial evaluation of a prototype accessibility information resource.

Chapter 6 presented the design, implementation and evaluation of the resource, called WebAIR. It addressed the research question: *What design attributes of an accessibility information resource are appropriate to support web developers in (a) the creation and evaluation of accessible websites; and (b) to foster a greater knowledge and understanding of web accessibility?* An initial evaluation with 26 professional web developers and 7 novice web developers elicited generally favourable first impressions of WebAIR. The chapter concluded with an account of further refinements made to the resource in response to the gathered feedback.

The third stage—**evaluation**—evaluated the ease of use and effectiveness of the previously developed accessibility information resource in supporting web developers to create and evaluate accessible websites.

The first study within this stage, presented in Chapter 7, was a laboratory-based evaluation of the resource with 48 novice web developers. This addressed the research question: *How effective is the resource in supporting novice web developers to confidently identify accessibility problems and solutions in existing websites?* The results of this study were disappointing: despite participants providing favourable ratings of the resource, WebAIR appeared to be no more effective (but, equally, no less effective) than WCAG 2.0. Although the study had several methodological limitations that may have confounded the results, it nevertheless raised the possibility that resources such as WebAIR may be less effective for newcomers to accessibility.

Addressing many of the methodological limitations in the previous study, the following study, presented in Chapter 8, was a real-world evaluation of WebAIR with 32 professional web developers. This addressed the research question: *How effective is the resource in supporting professional web developers to identify accessibility problems and solutions in a purpose-built website?* In addition to, once again, being rated favourably by participants, WebAIR appeared to be significantly more effective than WCAG 2.0. The study highlighted some of the strategies that web developers adopt when using accessibility information resources. It also demonstrated several weaknesses of WebAIR, particularly in regard to limited support for web developers who are already familiar with web accessibility. For typical web developers, however, who may have only a passing knowledge of the subject, WebAIR appeared to be a more useful, pragmatic accessibility information resource.

The final study, both of this stage, and of the programme of research, was an "in the wild" evaluation of the resource, presented in Chapter 9, with 15 professional web developers. The study marked a final opportunity to gain a more detailed understanding of how the resource is used by web developers. It addressed the research question: *How effective is the resource in supporting professional web developers to create an accessible website?* Despite a disappointing setback that resulted in the initial study being abandoned, each participant used WebAIR to create a (mainly) accessible webpage and provided valuable feedback on their experience. The difficulties I faced in running the initial "in the wild" evaluation indirectly addressed the research question: *How easily can the resource be integrated into professional web developers' existing working practices?* Even the most willing web developers faced numerous organisational constraints that prevented them from using WebAIR in their day-to-day web development activities. The generally positive feedback about WebAIR from participants in this study also went some way to addressing the research question: *Does exposure to the resource have any impact upon web developers' knowledge and understanding of web accessibility?*

My ultimate objective in completing this programme of research was to foster a greater understanding of web accessibility among web developers. Although I cannot claim to have had any widespread impact, I feel that my engagement with web developers, both through these studies and via presentations and publications made during the investigation, has, at least, gone some way to raising the profile and importance of web accessibility among this critical stakeholder group.

10.2. Thesis contributions

This thesis makes six distinct contributions to the field of web accessibility, relating to the three stages of the programme of research.

- The *analysis* stage contributes a systematic review of the web accessibility evaluation literature; an empirical demonstration of the organisational and procedural barriers that prevent the creation of accessible websites; and an empirical elicitation of web developers' mental models of web accessibility.
- The *intervention* stage contributes a definition of the necessary characteristics and attributes of resources to support web developers; and an embodiment of these findings in the development of a prototype accessibility information resource called WebAIR.
- The *evaluation* stage contributes a comprehensive evaluation of WebAIR's ease of use and effectiveness in supporting web developers to create and evaluate accessible websites, and to better integrate web accessibility into their existing workflows.

One contribution of this thesis is a systematic review of 397 web accessibility evaluation studies published over a 15-year period between 1999 and 2014. While many existing studies have compared the accessibility of a group of websites either at particular points in time (e.g. Sullivan & Matson, 2000; Lazar et al., 2003; Olalere & Lazar, 2011) or over a specified period of time (e.g. Hackett et al., 2005; Lazar & Greenidge, 2006; Loiacono et al., 2009; Hanson & Richards, 2013), none, to my knowledge, have appraised the methodology or synthesised the outcomes from such a large corpus of publications over such a large period of time. Although the considerable degree of variability and noise in the data precludes a more precise analysis, the research presented in Chapter 3 demonstrates a persistent occurrence of accessibility problems over the last two decades and highlights an alarming lack of methodological consistency in the literature.

Another contribution of the thesis is an empirical demonstration that web developers are hindered, not by limited awareness or concern, but by a lack of knowledge and practical guidance on how to make websites accessible. Existing studies (e.g. Lazar et al., 2004; Yesilada et al., 2015) have engaged web developers on the subject of web accessibility but have tended to do so via large-scale, anonymous surveys that are particularly vulnerable to social desirability bias (Nederhof, 1985). Although such surveys have monitored attitudinal trends towards web accessibility among web developers (e.g. Lazar et al., 2004; Farrelly, 2011), and have elicited their opinions on the causes of poor web accessibility (e.g. Freire et al., 2008, Olalere and Lazar, 2011; Putnam et al., 2012), few have managed to uncover the necessary level of detail to address the problem and to properly support web developers. The contextual inquiry, presented in Chapter 4, is, as far as I am aware, the first study to examine web developers in-depth in the context of their working environments. With social desirability bias being much easier to detect and mitigate during a contextual interview, the outcome is a more accurate portrayal of the organisational and procedural barriers that web developers face in creating and evaluating accessible websites.

Another contribution of this thesis is an empirical elicitation of web developers' mental models of web accessibility, which are revealed to be very limited and based on a faulty conceptual model that prioritises technical requirements over user needs and standards conformance over user experience. This presents a complex challenge for the development of accessibility information resources, which need to promote a user-centred conceptual model of web accessibility while respecting web developers' existing working practices and organisational constraints. The study presented in Chapter 5 is, to my knowledge, the first to explore the nature of web developers' confusion and uncertainty over web accessibility through the theoretical lens of mental models. While mental models remain an elusive concept, it nevertheless provides a useful framework for exploring the knowledge gaps, inconsistencies and uncertainties that result in inaccessible websites.

A further contribution of this thesis is a definition of the necessary characteristics and attributes of accessibility information resources to support web developers in the creation and evaluation of accessible websites. Key requirements, presented in Chapter 6, include:

- resources must speak the language of web developers, to avoid seeming vague, unspecific and reliant upon domain-specific terminology;
- the organisation and structure of resources must reflect how web developers approach the creation of websites, to avoid making the guidance appear abstract, arbitrary and unrelated to either the developers' work or the people they are intended to benefit;
- resources must present manageable amount of accessibility information and a realistic number of items for web developers to test, to avoid information and procedural overload; and
- resources must promote a user-centred conceptual model of web accessibility, portraying it as a human, rather than technical, challenge, and prioritising standards conformance over user experience.

Having defined the necessary characteristics and attributes of resources to support web developers, a further contribution of this thesis is an embodiment of the findings in the development of a prototype accessibility information resource, presented in Chapter 6, called WebAIR. Key features of the resource include:

- phrasing accessibility issues in WebAIR as concrete, objective questions for web developers to use in checking their work;
- organising the questions in WebAIR according to different types of web content to provide web developers with a concrete categorisation of information that they can more easily apply to their work;
- presenting only essential information about each issue and reducing the number of possible solutions to avoid information and procedural overload;

- providing web developers with just-in-time training in web accessibility concepts, allowing them to gradually learn about the domain of web accessibility while undertaking testing; and
- avoiding the concept of conformance or priority levels and including the justification or rationale behind each question to promote a more user-centred conceptual model of web accessibility.

A further contribution of this thesis is a comprehensive evaluation of WebAIR that demonstrates its ease of use and effectiveness in supporting web developers to create and evaluate accessible websites, and to better integrate web accessibility into their existing workflows. This research demonstrates that the design decisions made in the development of WebAIR have been largely successful. Professional and novice web developers rated the resource consistently and significantly more highly than WCAG 2.0 across a range of characteristics, suggesting it has strong face validity. The initial evaluation of WebAIR, presented in Chapter 6, indicated that such resources must be pitched at an appropriate level, particularly in terms of its language and complexity, to stimulate engagement from web developers. The laboratory evaluation of WebAIR, presented in Chapter 7, suggested that it may be less effective for web accessibility newcomers, due to a basic knowledge of the subject being necessary to navigate the resource. Conversely, the "real-world" evaluation of WebAIR, presented in Chapter 8, and the "in the wild" evaluation of WebAIR, presented in Chapter 9, suggested that WebAIR may be equally less effective for experienced web developers who are already well-versed in web accessibility, due to its occasional lack of detail and clarity. The web developers who are most likely to benefit from WebAIR are those who inspired this investigation: the web developers who are interested in and aware of web accessibility but who lack the knowledge and practical guidance to implement it. For them, the advantages that WebAIR presents, in terms of understanding and applying the information, makes it a more practical, pragmatic resource for creating and evaluating accessible websites.

10.3. Thesis limitations

The findings of this programme of research are subject to the limitations of the participant selection, the different methodologies and the measures used in each study and, of course, the quantitative and qualitative analysis applied to the results.

Despite my best efforts to recruit as widely as possible, many of the professional web developers who participated in this investigation had a similar background and level of experience. Also, because many of the participants in the earlier studies were recruited via i2Web consortium organisations, they tended to be more engaged towards web accessibility.

Publication and selection biases threatened the validity of the systematic review of the web accessibility evaluation literature, presented in Chapter 3, particularly in the exclusion of so-called “grey literature” and unpublished results. This threat was mitigated, however, through the selection of high quality, relevant research from a broad range of sources. Another concern was the considerable number of publications that were missed in our initial searches of the literature. However, many of these were picked up in the additional checks of each author’s publication record. The biggest limitation of this study was the considerable degree of variability and noise in the data that precluded a more precise analysis. It is only through continued efforts to unify and standardise the methodology and reporting of web accessibility evaluation studies that this limitation will be overcome.

Difficulties in recruiting participants who were willing (or permitted) to be closely observed at work threatened the success of the contextual inquiry investigation, presented in Chapter 4. Although recruiting many participants from i2Web consortium organisations proved convenient, it may have compromised the representativeness of the sample and the generalisability of the results. Furthermore, the limited amount of time available to interview each participant forced me to use a more directed approach to contextual inquiry. Although this was in keeping with the contextual inquiry methodology outlined by Beyer and Holtzblatt (1997) and provided additional benefits in terms of focusing the research, it inevitably created a more artificial situation. With regard to sample size, Beyer and Holtzblatt (1997) recommend interviewing between 10 and 20 people “unless the focus is very narrow” (p. 76). They also suggest recruiting participants from at least four to six organisations “to see variety” (p. 76). This study involved 13 professional web developers from eight organisations, which conforms to these recommendations, but nevertheless, may not be sufficient to draw strong conclusions. However, any threats to validity posed by this small sample size must be considered against the logistical challenge of managing the large amount of rich data generated by contextual interviews.

Concerns over the representativeness and generalisability of previous study were allayed by the interview study, presented in Chapter 5, which reaffirmed the results with a different and larger population. The limitations of this study stemmed from the ambiguous and intangible nature of mental models, which proved difficult to apply to the abstract concept of web accessibility. A triangulation of methods helped to address this ambiguity but, even so, it remains uncertain whether the study accurately and fully elicited web developers' mental models of web accessibility. Nevertheless, the study was successful in highlighting needs and requirements of web developers with regards to accessibility support.

Time limitations compromised the validity of the initial evaluation of WebAIR, presented in Chapter 6. By conducting this and the previous study in the same session, the participants may have unintentionally been fatigued. Furthermore, the limited amount of time that participants were exposed to each resource prevented anything more than a superficial inspection. I was also concerned that my known involvement in the design of WebAIR may have led to social desirability bias (Nederhof, 1985), with participants being reluctant to criticise the resource. Although impossible to completely eliminate such bias, the assistance of two other researchers during the data collection and my own efforts to distance myself from WebAIR when presenting the resource to participants hopefully mitigated its effect.

The laboratory-based evaluation of WebAIR, presented in Chapter 7, introduced a greater degree of rigour but still had a number of methodological limitations. Despite the potential educational incentive for taking part in the study, the student participants were not particularly motivated to take part. This may have led to them not putting much effort into the exercise, thus reducing the validity of the results. Efforts to provide a useful learning experience by having participants answer questions about each website may have also threatened the ecological validity of the study. Similarly, the amount of time participants had to complete each of the two evaluation tasks was not commensurate with that of a typical accessibility evaluation. The use of live websites in the accessibility evaluation tasks provided more ecologically valid stimuli but meant that the websites were uncontrollable and subject to unannounced updates and changes.

The "real-world" evaluation of WebAIR, presented in Chapter 8, addressed several limitations of the previous study, offering a more ecologically valid appraisal of the resource. Participants, who were professional web developers, evaluated a purpose-built website using their own web development environment, in their own working environment. Participants were more motivated to take part, which was bolstered by modest monetary compensation. Participants also had a longer amount of time in which to complete the evaluation task, although this was still much shorter than that of a typical accessibility evaluation. While these methodological adjustments improved the ecological validity of the study, some aspects, such as the use of a purpose-built website, remained artificial.

The biggest limitation of either iteration of the "in the wild" evaluation of WebAIR, presented in Chapter 9, was the quasi-experimental one-group case study design, in which only a single group of participants were involved. The lack of a comparison or control group of participants who did not use WebAIR (or who used an alternative accessibility information resource, such as WCAG 2.0) made it impossible to draw firm conclusions about WebAIR's effectiveness. That said, given the difficulties in recruiting participants for this study, as well as the limited time available in which to complete the study, a more rigorous experimental design would have been difficult to implement. Any methodological weaknesses that threaten the study's internal validity were offset by its high ecological validity. Participants completed the study under realistic conditions: using their own web development environment, in their own working environment, and in their own time.

10.4. Future work

Despite the significant role that web developers play in the creation and evaluation of accessible websites, surprisingly few studies have examined this population in any detail. This programme of research is a first attempt to study web developers in-depth in the context of their working environments to investigate why and how they struggle to create accessible websites. The investigation is one of many more that are necessary, however, to better understand and support this population and to foster a greater knowledge and awareness of web accessibility.

Specifically, future work should investigate a broader cross-section of web developers in terms of background, experience, and familiarity with web accessibility. As acknowledged in Section 10.3, many of the participants in this investigation had a similar background and level of experience and were already engaged towards web accessibility. It would be interesting, therefore, to engage with web developers who have little to no awareness or interest in web accessibility (a population that, I suspect, is far larger than imagined) to determine how best to support them. Similarly, non-professional hobbyist web developers who perhaps lack a technical background in developing websites would make an interesting comparison group.

This programme of research demonstrates that crucial to supporting web developers in making accessible websites is the need to comprehensively understand both how they think and operate. The contextual inquiry, presented in Chapter 4, and the "real-world" evaluation of WebAIR, presented in Chapter 8, gave practical insight into how web developers approach their work and the strategies they adopt in creating and evaluating accessible websites. In contrast to this, the interview study, presented in Chapter 5, examined what web developers understand about web accessibility and how this guides their approach. These different perspectives together contributed to the design of a usable and effective resource for web developers. Future research should continue to investigate this population in detail, both procedurally, through contextual exploration of their working practices, and cognitively, through exploration of their knowledge, comprehension and mental models of web accessibility.

While this research has demonstrated the effectiveness of ethnographic, user-centred research methods, further research is necessary to refine these methods to ensure they are appropriate to the target population and accurately reflect the phenomena under investigation. This is particular pertinent to the contextual design methodology, which, as described in Chapter 4, manifested several limitations when applied to the non-traditional 'digital workplaces' and working practices of web developers. While the adoption of a more interruptive approach to contextual inquiry, and the focus upon 'virtual' as well as 'physical' artefacts helped to mitigate the challenges, further investigations of this nature would benefit from efforts to define and refine methods for collecting, analysing and representing non-traditional, computer-based work. Similar efforts are necessary to improve the elicitation and representation of mental models, particularly for abstract concepts such as web accessibility.

Despite the apparent usability and effectiveness of WebAIR, further research is necessary to investigate the necessary characteristics of resources to better support web developers. In addition to designing and evaluating new and innovative ways of delivering accessibility information, research should continue to reflect on and evaluate existing solutions. This programme of research demonstrates that web developers undoubtedly struggle with existing tools, guidelines and resources and, while we must not dismiss the considerable amount of work that has culminated in these solutions, it is important to critically appraise what is working and what is not. Specifically, future work should explore the usability of resources such as WCAG 2.0 to determine how well they serve not only web developers, but also other stakeholders involved in the value chain of accessible web development (see Section 1.3). It is particularly important to explore how well such resources support the learning and remembering of accessibility principles, as opposed to providing ready-made solutions that web developers can apply but not necessarily understand.

Previous studies have tended to evaluate resources such as WCAG 2.0 under somewhat artificial experimental conditions. Researchers typically conduct studies of this nature in university classrooms and laboratories using naive participants undertaking unrepresentative tasks and using unrepresentative materials. Although this approach provides a degree of experimental rigour and control, it lacks the ecological validity necessary to determine the resources' real-world effectiveness. This thesis presents a series of evaluations conducted with both novice and professional web developers and under both laboratory and real-world conditions. With each evaluation of WebAIR, I attempted to increase the ecological validity, culminating in an "in the wild" evaluation. In my opinion, the use of professional web developers, using their own web development environment, in their own working environment, and in their own time, served to greatly enhance the relevance and generalisability of my results. Further research should, where possible, aim to conduct evaluations under ecologically valid experimental conditions to produce relevant, useful results that may be generalised to the broader web developer community.

10.5. The future of WebAIR

This programme of research has demonstrated the usability and effectiveness of WebAIR in supporting web developers to create and evaluate accessible websites. While the current format and presentation of WebAIR (a standalone multi-page website) appears effective, there is much potential for a more innovative delivery method. Key to web developers' adoption of the resource is its integration into their existing working practices. This could be encouraged by incorporating the resource into the IDEs and code editors that web developers commonly use, such as Visual Studio[71], Atom[72] or Brackets[73]. WebAIR's integration into the i2Web EASI tool demonstrated the potential of this approach. WebAIR could also exist as a browser extension, which web developers could query when previewing their websites. Such initiatives would make the resource more readily available during the coding stage of web development but, as cautioned in Chapter 6, efforts to automate and streamline accessibility evaluation must not come at the expense of educating and informing web developers.

The suggestions above are based on the assumption that WebAIR will remain a static resource. A more dynamic, interactive approach could allow web developers to record their responses to the questions in WebAIR, affording opportunities to benchmark and monitor the accessibility of websites. To further streamline the development process, the questions in WebAIR could be filtered according to different criteria and contexts, such as the type of website under evaluation, the type of content under evaluation, or the type of web development role (e.g. front-end or back-end development). WebAIR could even be adapted to support other stakeholders (e.g. graphic designers, content creators etc.) in the value chain of accessible web development. The W3C's How to Meet WCAG 2.0 tool[74] takes a similar approach, allowing users to define a customisable subsection of the guidelines. Participants in several of the evaluations also suggested incorporating interactive "code playgrounds" (similar to Codepen[75], JSFiddle[76], JS Bin[77] etc.) into WebAIR's 'More Information' pages, which would allow web developers to rapidly create, edit, preview and share accessible code samples.

A persistent threat to the accuracy and relevance of resources such as WebAIR or WCAG 2.0 is the rapid progress of web technologies as well as the emergence and refinement of accessible web techniques. Although the recently rechartered (and renamed) W3C Accessibility Guidelines Working Group, which is responsible for WCAG 2.0 and associated resources, states a commitment to "produce updated guidance for accessibility on a regular interval" (Kirkpatrick, O Connor, & Cooper,

---

71 https://www.visualstudio.com
72 https://atom.io
73 http://brackets.io
74 https://www.w3.org/WAI/WCAG20/quickref
75 https://codepen.io
76 https://jsfiddle.net
77 https://jsbin.com

2017a), it is nevertheless restricted by the W3C's lengthy review process (McCathie Nevile, 2017b). To address this issue in WebAIR, one participant suggested it should be made open source and should function similar to a wiki, allowing members of the web development and web accessibility communities to directly edit and contribute. While this would impose obvious risks regarding the accuracy and authority of the information, it would nevertheless ensure the content remains relevant and up-to-date, as well as engaging web developers in the accessible web development process.

10.6. Developments since this research was conducted

Since this programme of research was conducted, numerous other resources have been developed that share some of the characteristics of WebAIR.

For example, in 2015, a digital services agency within the US federal government, called 18F, launched an Accessibility Guide[78] to provide a repository for best practices. Similar to WebAIR, the guide is organised according to easily applicable categories (such as images, forms, multimedia). However, whereas WebAIR's categories correspond directly to different types of web content, the 18F Accessibility Guide also includes categories relating to different behaviours (e.g. flashing content, hidden content) and technical compatibility (e.g. keyboard access, CSS dependence). As in WebAIR, each section of the 18F Accessibility Guide includes a brief description of the issue, written in developer-friendly language, as well as steps on how to test it. Notably, the descriptions rarely include a rationale or justification but instead focus on objective accessibility criteria. Like WebAIR, code examples (of both passes and fails) are provided as well as links to additional resources. The guide also includes a "checklist" for web developers, which draws together the different sections but, unlike WebAIR, it prioritises the advice according to its potential impact on disabled users. The 18F Accessibility Guide is open source and hosted on GitHub, with users encouraged to submit issues and pull requests.

In 2015, the French government undertook a major revision of its national standard for digital accessibility in the public sector, the Référentiel Général d'Accessibilité pour les Administrations (RGAA). Incorporating the AccessiWeb framework for testing conformity to WCAG 2.0 success criteria, developed by the BrailleNet association,

[78] https://accessibility.18f.gov/

RGAA 3.0[79] constitutes a reference system that proposes criteria and tests to verify conformance to WCAG 2.0. Similar to WebAIR, RGAA is organised by different content types (e.g. images, tables, links etc.), which serves to grounds the domain-specific groupings of WCAG 2.0. Another similarity to WebAIR is how each issue in RGAA is phrased as a question (e.g. *Does each image have a text alternative? Does each complex data table have a summary?*) As in WebAIR, the questions are mapped to WCAG 2.0 success criteria and techniques. However, while the questions in RGAA include links to a glossary of terms, no additional information is provided to justify or explain each issue. Also, due to RGAA's role as a technical standard, it uses very precise but verbose language that incorporates much of the jargon and terminology found in WCAG 2.0. While the RGAA remains the property of the State of France, it is published under an open licence, making it freely reusable. RGAA provides an objective set of tests that eliminate much of the ambiguity and subjectivity found in WCAG 2.0 but offers little opportunity to further web developers' knowledge of web accessibility.

The A11Y Project[80], conceived in 2012, is an open-source community-driven resource for front-end web developers that aims to make web accessibility easier to implement. It includes a collection of articles, written in developer-oriented language, that are themed around different aspects of web accessibility (e.g. how to use skip navigation links; getting started with ARIA; what is colour contrast?). An accessible widget and pattern library is also provided with links to code examples. While the resource could benefit from a more coherent structure, it includes a Web Accessibility Checklist, which, similar to WebAIR, provides a unified list of accessibility issues for web developers to consider. As in WebAIR, the checklist is organised by different content types (e.g. links, images, forms) but no additional information is provided to justify or explain each point. The A11Y Project is an incomplete work-in-progress but is regularly updated and benefits from strong engagement from the web accessibility community. It is also open source and hosted on GitHub.

In a similar vein to The A11Y project, The A11Y Style Guide[81], conceived in 2017, presents a collection of accessible components for creating accessible webpages, as well as links to related tools, articles, and WCAG guidelines. Although some of the style

[79] http://disic.github.io/rgaa_referentiel_en/criteria.html (English translation)
[80] http://a11yproject.com/
[81] http://a11y-style-guide.com/style-guide/

guide is, like WebAIR, organised according to different content types (e.g. forms, media, navigation etc.), topics such as headings and tables are grouped under "Structure", whereas miscellaneous topics such as buttons and colours are grouped under "General". While a more intuitive organisation would perhaps make the resource easier to navigate, the current small number of categories (seven, plus an overview) ensures the resource remains manageable. A brief description, written in developer-oriented language, is included in each topic, along with code examples, WCAG 2.0 references, and links to additional resources. Although some justification is included, the focus of the A11Y Style Guide is very much on providing practical guidance. It is open source and hosted on GitHub, with users encouraged to contribute or build upon it.

The existence of these resources serves to validate many of the design decisions made in the development of WebAIR. It also reaffirms the difficulty that web developers currently face in creating and evaluating accessible websites and the need for usable resources that provide access to relatable, relevant and rigorous information.

10.7. Overall conclusions

The programme of research has demonstrated the effectiveness of ethnographic, user-centred research methods in collecting and analysing data about web developers and proposing design solutions to meet their needs. It has contributed to a better understanding of web developers' current working practices and how they impact on web developers' knowledge and understanding of web accessibility. It has fostered a greater awareness of the necessary characteristics and attributes of effective accessibility information resources and has proposed a solution that embodies those findings. It has also evaluated the effectiveness of the proposed solution in supporting web developers to create and evaluate accessible websites, and to better integrate web accessibility into their existing workflows.

Over the seven-year course of this research, two important caveats have become apparent. Firstly, the support that resources such as WebAIR may provide to web developers is only one part of the solution. The initial study of this thesis highlighted the sustained poor levels of accessibility in websites from around the world. For these to improve, we still need to shift web developers away from their current development-centred conceptual model and towards a more humanistic, user-centred conceptual model. Despite broad approval for WebAIR from the web developers who participated in this research, it may not necessarily be the most effective way of changing attitudes towards and fostering awareness of web accessibility. In addition to tools, guidelines and resources, the needs of disabled people must be investigated by other means, such as user testing and field research. It is crucial that web developers do not become too reliant upon tools, guidelines and resources at the expense of acknowledging the actual experiences and perspectives of disabled people.

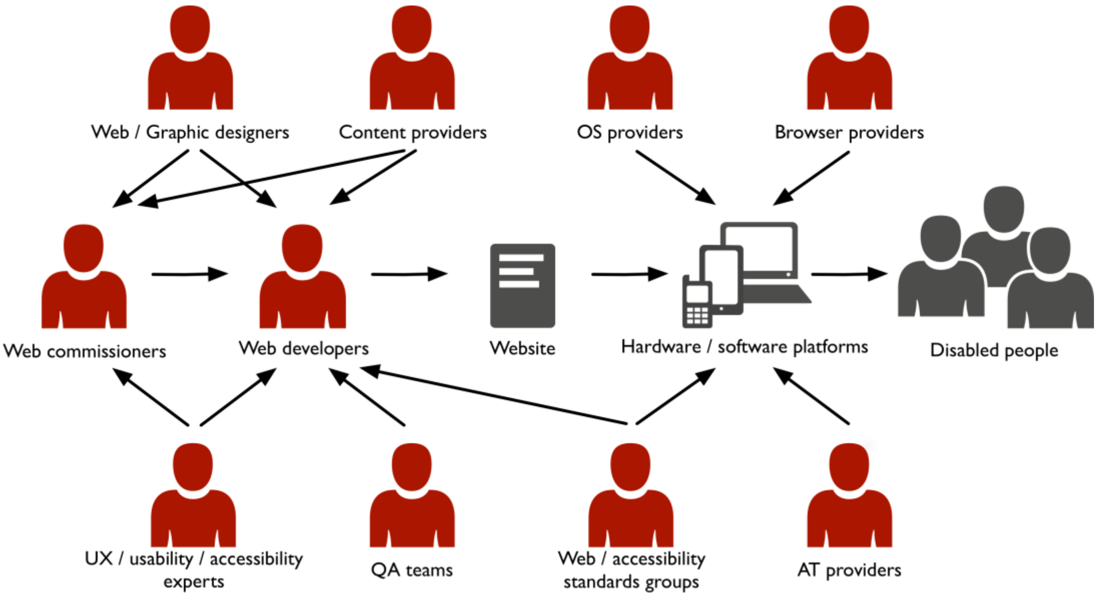


Figure 10.1: The shared responsibility of stakeholders involved in the value chain of accessible web development

Secondly, the difficulties that web developers encounter in creating and evaluating accessible websites is only one part of the problem. While I still maintain that web developers are the stakeholders with the greatest responsibility for web accessibility, this programme of research has demonstrated that we must broaden our scope to include more than just web developers (see Figure 10.1). The final study of this thesis, in particular, highlighted that even the most willing web developers face numerous organisational constraints that prevent them incorporating accessibility into their day-to-day web development activities. It is clear that web accessibility needs to become part of a broader organisational approach and must be integrated into all stages of design and development. Crucially, it needs to involve all stakeholders involves in the value chain of accessible web development.

A web that is *for everyone* must be *everyone's responsibility*.

References

[Note: References from the systematic review of web accessibility evaluation literature, presented in Chapter 3, are included in Appendix G]

Abou-Zahra, S. (Ed.) (2007). Evaluation and Report Language (EARL) 1.0 Schema - W3C Working Draft 23 March 2007. Retrieved October 31, 2017 from https://www.w3.org/TR/2007/WD-EARL10-Schema-20070323/

Adams-Spink, G. (2004). Websites 'failing' disabled users. *BBC News*. Retrieved October 31, 2017 from http://news.bbc.co.uk/1/hi/technology/3626533.stm

Adams-Spink, G. (2006). New standards for website access. *BBC News*. Retrieved October 31, 2017 from http://news.bbc.co.uk/1/hi/technology/4783686.stm

Adams-Spink, G. (2008). New guidelines boost web access. *BBC News*. Retrieved October 31, 2017 from http://news.bbc.co.uk/1/hi/technology/7789622.stm

Aizpurua, A., Arrue, M., Harper, S., & Vigo, M. (2014). Are users the gold standard for accessibility evaluation?. In *Proceedings of the 11th Web for All Conference (W4A 2014)*. New York, NY, USA: ACM.

Allan, J., Lowney, G., Patch, K., & Spellman, J. (Eds.) (2015). User Agent Accessibility Guidelines (UAAG) 2.0 - W3C Working Group Note 15 December 2015. Retrieved October 31, 2017 from https://www.w3.org/TR/UAAG20/

Allen, R.B. (1997) Mental Models and User Models. In M. G. Helander, T. K. Landauer, & P. V. Prabhu (Eds.), *Handbook of Human-Computer Interaction, Volume 1* (pp.49-63). Amsterdam, The Netherlands, Elsevier Science.

Alonso, F., Fuertes, J. L., González, Á. L., & Martínez, L. (2010a). Alonso F., Fuertes J.L., González Á.L., Martínez L. (2010) Evaluating Conformance to WCAG 2.0: Open Challenges. In Miesenberger K., Klaus J., Zagler W., Karshmer A. (Eds.), *Computers Helping People with Special Needs. ICCHP 2010. Lecture Notes in Computer Science, vol 6179*. Berlin/Heidelberg, Germany: Springer.

Alonso, F., Fuertes, J. L., González, Á. L., & Martínez, L. (2010b). On the testability of WCAG 2.0 for beginners. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A 2010)*. New York, NY, USA: ACM.

Apple Computer Inc. (2017). Apple Human Interface Guidelines. Retrieved October 31, 2017 from https://developer.apple.com/library/content/documentation/UserExperience/Conceptual/OSXHIGuidelines/DesignPrinciples.html

Australian Government. (1992). Australian Disability Discrimination Act. Canberra: Australian Government Printing Services. Retrieved October 31, 2017 from https://www.legislation.gov.au/Details/C2017C00339

Bailey, C., & 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 2010)*. New York, NY, USA: ACM.

Bailey, C., & Pearson, E. (2011). Development and trial of an educational tool to support the accessibility evaluation process. In *Proceedings of the 2011 International Cross Disciplinary Conference on Web Accessibility (W4A 2011)*. New York, NY, USA: ACM.

Bailey, J., & Burd, E. (2005). Tree-map visualisation for web accessibility. In *Proceedings of the Computer Software and Applications Conference* (pp. 275-280). London, UK: IEEE.

Becker, S. A. (2004). E-government visual accessibility for older adult users. *Social Science Computer Review*, 22(1), 11-23.

Benavídez, C., Fuertes, J. L., Gutiérrez, E., & Martínez, L. (2006). Teaching web accessibility with “contramano” and hera. In Miesenberger K., Klaus J., Zagler W.L., Karshmer A.I. (Eds.) *Computers Helping People with Special Needs. ICCHP 2006. Lecture Notes in Computer Science, vol 4061*. Berlin/Heidelberg, Germany: Springer.

Beyer, H., & Holtzblatt, K. (1997). *Contextual Design: Defining Customer-Centered Systems.* San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Bibby, P.A. (1992). Mental Models, Instructions and Internalization. In Rogers, Y., Rutherford, A. and Bibby, P.A. (Eds.) *Models in the Mind: Theory, Perspective and Applications* (pp.153-172). London, UK: Academic Press.

Bigham, J. P., & Ladner, R. E. (2007). Accessmonkey: a collaborative scripting framework for web users and developers. In *Proceedings of the 2007 International Cross Disciplinary Conference on Web Accessibility (W4A 2007)* (pp. 25-34). New York, NY, USA: ACM.

Bigham, J. P., Brudvik, J. T., & Zang, B. (2010). Accessibility by demonstration: enabling end users to guide developers to web accessibility solutions. In *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2010)* (pp. 35-42). New York, NY, USA: ACM.

Bigham, J. P., Prince, C. M., & Ladner, R. E. (2008). WebAnywhere: a screen reader on-the-go. In *Proceedings of the 2008 International Cross Disciplinary Conference on Web Accessibility (W4A 2008)* (pp. 73-82). New York, NY, USA: ACM.

Brajnik, G. (2006). Web Accessibility Testing: When the Method Is the Culprit. In Miesenberger K., Klaus J., Zagler W.L., Karshmer A.I. (Eds.) *Computers Helping People with Special Needs. ICCHP 2006. Lecture Notes in Computer Science, vol 4061*. Berlin/Heidelberg, Germany: Springer.

Brajnik, G. (2008). A comparative test of web accessibility evaluation methods. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2008)* (pp. 113-120). New York, NY, USA: ACM.

Brajnik, G. (2009). Validity and reliability of web accessibility guidelines. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2009)* (pp. 131-138). New York, NY, USA: ACM.

Brajnik, G., & Lomuscio, R. (2007). SAMBA: a semi-automatic method for measuring barriers of accessibility. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2007)* (pp. 43-50). New York, NY, USA: ACM.

Brewer, J. (Ed.) (1998). Web Accessibility Initiative Education & Outreach Working Group Charter. Retrieved October 31, 2017 from https://www.w3.org/WAI/EO/charter

Caldwell, B., Cooper, M., Reid, L. G., & Vanderheiden, G. (Eds.) (2008a). Web Content Accessibility Guidelines 2.0 - W3C Recommendation 11 December 2008. Retrieved October 31, 2017 from https://www.w3.org/TR/WCAG20/

Caldwell, B., Cooper, M., Reid, L. G., & Vanderheiden, G. (Eds.) (2008b). Techniques for WCAG 2.0 – W3C Working Group Note 11 December 2008. Retrieved October 31, 2017 from: https://www.w3.org/TR/2008/NOTE-WCAG20-TECHS-20081211/

Caldwell, B., Cooper, M., Reid, L. G., & Vanderheiden, G. (Eds.) (2008c). Understanding WCAG 2.0 - W3C Working Group Note 11 December 2008. Retrieved October 31, 2017 from: https://www.w3.org/TR/2008/NOTE-UNDERSTANDING-WCAG20-20081211/

Calvo, R., Iglesias, A., & Moreno, L. (2012). Is Moodle accessible for visually impaired people?. In Filipe J., Cordeiro J. (Eds.) *Web Information Systems and Technologies. WEBIST 2011. Lecture Notes in Business Information Processing, vol 101.* Berlin/Heidelberg, Germany: Springer.

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, 42(3), 294-320.

Chen, Y., Chen, Y., & Shao, M. (2006). 2005 Accessibility diagnosis on the government web sites in Taiwan, R.O.C. In *Proceedings of the International Cross Disciplinary Workshop on Web Accessibility (W4A 2006)*. New York, NY, USA: ACM.

Chevalier, A., Fouquereau, N., & Vanderdonckt, J. (2009). The influence of a knowledge-based system on designers' cognitive activities: a study involving professional web designers. *Behaviour & Information Technology*, 28(1), 45-62.

Chisholm, W., Vanderheiden, G., & Jacobs, I. (Eds.) (1999). Web Content Accessibility Guidelines 1.0 – W3C Recommendation 5-May-1999. Retrieved October 31, 2017 from https://www.w3.org/TR/WCAG10/

Chisholm, W., Vanderheiden, G., & Jacobs, I. (Eds.) (2000). Techniques for Web Content Accessibility Guidelines 1.0 – W3C Note 6 November 2000. Retrieved October 31, 2017 from https://www.w3.org/TR/WAI-WEBCONTENT-TECHS/

Colwell, C., & Petrie, H. (2001). Evaluation of guidelines for designing accessible web content. *ACM SIGCAPH Computers and the Physically Handicapped*, 70, 11-13.

Comeaux, D., & Schmetzke, A. (2013). Accessibility of academic library web sites in North America: Current status and trends (2002-2012). *Library Hi Tech*, 31(1), 8-33.

Conant, R. C., & Ross Ashby, W. (1970). Every good regulator of a system must be a model of that system. *International Journal of Systems Science*, 1(2), 89-97.

Correia, M., Cruz, G., Nunes, R., Martins, J., Gonçalves, R., Paredes, H., & Martins, P. (2013, July). Network for all: a proposal for an accessible social media aggregator solution. In Stephanidis C., Antona M. (Eds.) *Universal Access in Human-Computer Interaction. User and Context Diversity. UAHCI 2013. Lecture Notes in Computer Science, vol 8010*. Berlin/Heidelberg, Germany: Springer.

Craig, J., & Cooper, M. (Eds.) (2014). Accessible Rich Internet Applications (WAI-ARIA) 1.0. Retrieved October 31, 2017 from https://www.w3.org/TR/wai-aria/

Craik, K. J. W. (1943). *The nature of explanation*. Cambridge, UK: University Press.

DiLallo, J. J., & Siegfried, R. M. (2009). The accessibility of college and university home pages in the state of New York. *Journal of Information Systems Applied Research*, 2(4).

Disability Rights Commission. (2004). The Web: Access and Inclusion for Disabled People. London, UK: The Stationery Office.

Donnelly, A., & Magennis, M. (2002). Making accessibility guidelines usable. In Carbonell N., Stephanidis C. (Eds.) *Universal Access Theoretical Perspectives, Practice, and Experience. UI4ALL 2002. Lecture Notes in Computer Science, vol 2615*. Berlin/Heidelberg, Germany: Springer.

Eggert, E., & Abou-Zahra, S. (Eds.) (2016). Web Accessibility Evaluation Tools List. Retrieved October 23, 2017 from https://www.w3.org/WAI/ER/tools/

European Union, Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies, OJ L 327, 2.12.2016, p. 1-15.

Farrelly, G. (2011). Practitioner barriers to diffusion and implementation of web accessibility. *Technology and Disability*, 23 (4), 223-232.

Feigh, K. M. (2006). Contextual inquiry of a 100 aircraft regional airline systems operation center. Georgia Tech Cognitive Engineering Center, Technical Report GTCEC2006-0001.

Feingold, L. (2017). Digital Accessibility Laws Around The Globe. Retrieved 31 October, 2017 from http://www.lflegal.com/2013/05/gaad-legal/

Fischer, G. (1991). The Importance of Models in Making Complex Systems Comprehensible. In D. Ackerman, & M. Tauber (Eds.) *Mental Models and Human Computer Communication: Proceedings of the 8th Interdisciplinary Workshop on Informatics and Psychology), Volume 2* (pp.3-36). Amsterdam, The Netherlands: Elsevier Science.

Freire, A. P. (2012). Disabled people and the Web: User-based measurement of accessibility (Doctoral dissertation, University of York, UK).

Freire, A. P., de Mattos Fortes, R. P., Barroso Paiva, D. M., & Santos Turine, M. A. (2007). Using screen readers to reinforce web accessibility education. *ACM SIGCSE Bulletin*, 39 (3), 82-86.

Freire, A. P., Russo, C. M., & Fortes, R. P. M. (2008). A survey on the accessibility awareness of people involved in web development projects in Brazil. In *Proceedings of the 2008 International Cross Disciplinary Cconference on Web Accessibility (W4A 2008)* (pp. 87-96). New York, NY, USA: ACM.

Furniss, D., & Blandford, A. (2006). Understanding emergency medical dispatch in terms of distributed cognition: a case study. *Ergonomics*, 49(12-13), 1174-1203.

Gambino, O., Pirrone, R., & Di Giorgio, F. (2014). Accessibility of the Italian institutional web pages: a survey on the compliance of the Italian public administration web pages to the Stanca Act and its 22 technical requirements for web accessibility. *Universal Access in the Information Society*, pp. 1-8.

Garrett, J. J. (2005). Ajax: A New Approach to Web Applications. Archived from the original on 2 July 2008. Retrieved 31 October, 2017 from https://web.archive.org/web/20080702075113/http://www.adaptivepath.com/ideas/essays/archives/000385.php

Gellenbeck, E. (2005). Integrating accessibility into the computer science curriculum. *Journal of Computing Sciences in Colleges*, 21(1), 267-273.

Gentner, D., & Stevens, A.L. (Eds.). (1983). *Mental models*. Hillsdale, NJ, USA: Ertbaum.

Gibson, B. (2007). Enabling an accessible web 2.0. In *Proceedings of the 2007 International Cross Disciplinary Conference on Web Accessibility (W4A 2007)* (pp. 1-6). New York, NY, USA: ACM.

Gonçalves, R., Martins, J., Pereira, J., Oliveira, M. A. Y., & Ferreira, J. J. P. (2013). Enterprise Web accessibility levels amongst the Forbes 250: Where art thou o virtuous leader?. *Journal of Business Ethics*, 113(2), 363-375.

Government Digital Service (2017). Making your service accessible: an introduction. Retrieved October 31, 2017 from https://www.gov.uk/service-manual/helping-people-to-use-your-service/making-your-service-accessible-an-introduction

"Government sites fail web tests". (2005). *BBC News*. Retrieved October 31, 2017 from http://news.bbc.co.uk/1/hi/technology/4853000.stm

Greeff, M., & Kotzé, P. (2009). A lightweight methodology to improve web accessibility. In Proceedings of the 2009 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT 2009) (pp. 30-39). New York, NY, USA: ACM.

Gregor, P., Sloan, D., & Newell, A. F. (2005). Disability and technology: building barriers or creating opportunities?. *Advances in Computers*, 64, 283-346.

Hackett, S., & Parmanto, B. (2009). Homepage not enough when evaluating web site accessibility. *Internet Research*, 19(1), 78-87.

Hackett, S., Parmanto, B., & Zeng, X. (2005). A Retrospective Look at Website Accessibility over Time. *Behaviour and Information Technology*, 24(6), 407-417.

Hanson, V. L., & Richards, J. T. (2013). Progress on website accessibility? *ACM Transactions on the Web (TWEB)*, 7(1), 2.

Harper, S., & Chen, A. Q. (2012). Web accessibility guidelines. *World Wide Web*, 15(1), 61-88.

Harrison, S. M. (2005). Opening the eyes of those who can see to the world of those who can't: a case study. ACM SIGCSE Bulletin, 37(1), pp.22-26).

Hart, T. A., & Chaparro, B. (2004). Evaluation of websites for older adults: How "senior-friendly" are they. *Usability News*, 6(1), 12.

Hart, T. A., Chaparro, B. S., & Halcomb, C. G. (2004). Designing websites for older adults: The relationship between guideline compliance and usability. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 48, No. 2, pp. 271-274). Los Angeles, CA, USA: SAGE Publications.

Hendry, D. G., & Efthimiadis, E. N. (2008). Conceptual Models for Search Engines. In A. Spink & M. Zimmer (Eds.), *Web Search. Information Science and Knowledge Management, vol 14*. Berlin/Heidelberg, Germany: Springer.

Henry, S. L. (Ed.) (2005a). Introduction to Web Accessibility. Retrieved October 31, 2017 from https://www.w3.org/WAI/intro/accessibility.php

Henry, S. L. (Ed.) (2005b). Essential Components of Web Accessibility. Retrieved October 31, 2017 from https://www.w3.org/WAI/intro/components.php

Henry, S. L. (Ed.) (2010). Involving Users in Evaluating Web Accessibility. Retrieved October 31, 2017 from https://www.w3.org/WAI/eval/users.html

Henry, S. L. (Ed.) (2017a). Web Accessibility Initiative Education and Outreach Working Group Charter. Retrieved October 31, 2017 from https://www.w3.org/WAI/EO/charter2017

Henry, S. L. (Ed.) (2017b). WAI Resources. Retrieved October 31, 2017 from https://www.w3.org/WAI/Resources/Overview

Henry, S. L., & Brewer, J. (Eds.) (2000). WAI Mission and Organisation. Retrieved October 31, 2017 from https://www.w3.org/WAI/about.html

Henry, S. L., & Popolizio, P. (Eds.) (2001). Quick Tips to Make Accessible Web Sites. Retrieved October 31, 2017 from https://www.w3.org/WAI/quicktips/

Henry, S. L., Abou-Zahra, S., & White, K. (Eds.) (2016). Accessibility, Usability, and Inclusion: Related Aspects of a Web for All. Retrieved October 31, 2017 from https://www.w3.org/WAI/intro/usable

Hertz, H. R. (1899). *The Principles of Mechanics Presented in a New Form* (English translation), London, UK: Macmillan Press.

Hickson, I., Berjon, R., Faulkner, S., Leithead, T., Navara, E. D., O'Connor, E., & Pfeiffer, S. (Eds.) (2014). HTML5: A vocabulary and associated APIs for HTML and XHTML – W3C Recommendation 28 October 2014. Retrieved October 31, 2017 from https://www.w3.org/TR/html5/

Hong, S. G., Trimi, S., Kim, D. W., & Hyun, J. H. (2015). A Delphi Study of Factors Hindering Web Accessibility for Persons with Disabilities. *Journal of Computer Information Systems*, 55(4), 28-34.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277-1288.

International Business Machines Corporation. (1992). Object-oriented interface design: IBM Common User Access guidelines. Carmel, IN, USA: Que Corporation.

ISO/IEC (2012). ISO/IEC 40500:2012. Information technology: W3C Web Content Accessibility Guidelines (WCAG) 2.0. Retrieved October 31, 2017 from http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=58625

Jacobs, I., Gunderson, J., & Hansen, E. (Eds.) (2002). User Agent Accessibility Guidelines 1.0 – W3C Recommendation 17 December 2002. Retrieved October 23, 2017 from https://www.w3.org/TR/WAI-USERAGENT/

Jaeger, P. T. (2006). Assessing Section 508 compliance on federal e-government Web sites: A multi-method, user-centered evaluation of accessibility for persons with disabilities. *Government Information Quarterly*, 23(2), 169-190.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, UK: University Press.

Johnson-Laird, P. N. (1989). *Mental models*. In M. I. Posner (Ed.), Foundations of cognitive science. Cambridge, MA, USA: MIT Press.

Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Review of Psychology*, 50(1), 109-135.

Johnson-Laird, P.N. (2004) The history of mental models. In Manktelow, K., and Chung, M.C. (Eds.) *Psychology of Reasoning: Theoretical and Historical Perspectives* (pp.179-212). New York: Psychology Press.

Jones, N. A., Ross, H., Lynam, T., Perez, P., & Leitch, A. (2011). Mental models: an interdisciplinary synthesis of theory and methods. *Ecology and Society*, 16 (1), 46-46.

Kamoun, F., Al Mourad, B. M., & Bataineh, E. (2013). WCAG 1.0 versus WCAG 2.0 web accessibility compliance: A case study. In *Proceedings of the International Conference on Digital Information Processing, E-Business and Cloud Computing* (pp. 94-101).

Kapsi, M., Vlachogiannis, E., Darzentas, J. S., & Spyrou, T. (2009). A preliminary feedback for the WCAG 2.0-WCAG 1.0 Vs WCAG 2.0 evaluation study. In *Proceedings of the PETRA 2009 Conference.*

Kempton, W. (1986). Two theories of home heat control. *Cognitive Science*, 10(1), 75-90.

Kerr, S. T. (1990). Wayfinding in an electronic database: The relative importance of navigational cues vs. mental models. *Information Processing & Management*, 26(4), 511-523.

King, M., Nurthen, J., Bijl, M., & Cooper, M. (Eds.) (2017). WAI-ARIA Authoring Practices 1.1 – W3C Working Draft 28 June 2017. Retrieved October 31, 2017 from https://www.w3.org/TR/wai-aria-practices/

Kirkpatrick, A., O Connor, J., & Cooper, M. (2017a). Accessibility Guidelines Working Group Charter. Retrieved October 31, 2017 from: https://www.w3.org/2017/01/ag-charter

Kitchenham, B.A., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering, Technical Report EBSE-2007-01, School of Computer Science and Mathematics, Keele University.

Klein, D., Myhill, W., Hansen, L., Asby, G., Michaelson, S., & Blanck, P. (2003). Electronic doors to education: Study of high school website accessibility in Iowa. *Behavioral Sciences and the Law*, 21, 27–49.

Knight, J. (2003). Attitudes to Web Accessibility. Archived from the original on 11 November 2010. Retrieved 23 October, 2017 from https://web.archive.org/web/20101111042840/http://www.usabilitynews.com/news/article1321.asp

Korn, P. (Ed.) (2009). ACCESSIBLE Project: Deliverable 2.2b – User needs and System Requirements Specification. Retrieved October 31, 2017 from http://www.accessible-eu.org/index.php/deliverables.html

Krippendorff, K. (1995). On the Reliability of Unitizing Continuous Data. *Sociological Methodology*, 25. 47-76.

Kubische, L., Cullen, K., Dolphin, C., Laurin, S., & Cederbom, A. (Eds.) (2013). Study on Assessing and Promoting E-Accessibility. *European Commission*. Retrieved 31 October, 2017 from https://ec.europa.eu/digital-single-market/en/news/study-assessing-and-promoting-e-accessibility

LaToza, T. D., Venolia, G., & DeLine, R. (2006). Maintaining mental models: a study of developer work habits. In Proceedings of the 28th International Conference on Software Engineering (ICSE 2006) (pp. 492-501). New York, NY, USA: ACM.

Laurin, S., Cederbom, A., Martinez-Usero, J., Kubitschke, L., Naughton, C., Moledo, A., Shermer, R., & Abou-Zahra, S. (Eds.) (2015). Monitoring methodologies for web-accessibility in the European Union. *European Commission*. Retrieved 31 October, 2017 from https://ec.europa.eu/digital-single-market/en/news/monitoring-methodologies-web-accessibility-european-union

Lazar, J., & Greenidge, K. (2006). One year older, but not necessarily wiser: an evaluation of homepage accessibility problems over time. *Universal Access in the Information Society*, 4(4), 285-291.

Lazar, J., Beere, P., Greenidge, K., & Nagappa, Y. (2003). Web accessibility in the Mid-Atlantic United States: a study of 50 homepages. *Universal Access in the Information Society*, 2, 331–341.

Lazar, J., Dudley-Sponaugle, A., & Greenidge, K. (2004). Web accessibility: a study of webmaster perceptions. *Computers in Human Behaviour*, 20(2), 269-288.

Lazar, J., Wentz, B., Almalhem, A., Catinella, A., Antonescu, C., Aynbinder, Y., Bands, M., Bastress, E., Chan, B., Chelden, B., & Feustel, D. (2013). A longitudinal study of state government homepage accessibility in Maryland and the role of web page templates for improving accessibility. *Government Information Quarterly*, 30(3), 289-299.

Lee, S., Kim, B. G., & Kim, J. G. (2007). Accessibility evaluation of Korean e-government. In Stephanidis C. (Ed.) *Universal Access in Human-Computer Interaction. Applications and Services. UAHCI 2007. Lecture Notes in Computer Science, vol 4556*. Berlin/Heidelberg, Germany: Springer.

Leporini, B., Paternò, F., & Scorcia, A. (2006). Flexible tool support for accessibility evaluation. *Interacting with Computers*, 18(5), 869-890.

Littman, D. C., Pinto, J., Letovsky, S., & Soloway, E. (1987). Mental models and software maintenance. *Journal of Systems and Software*, 7(4), 341-355.

Loiacono, E. T., & McCoy, S. (2006). Website accessibility: a cross-sector comparison. *Universal Access in the Information Society*, 4(4), 393-399.

Loiacono, E. T., Romano, N. C., & McCoy, S. (2009). The state of corporate website accessibility. *Communications of the ACM*, 52(9), 128-132.

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human communication research*, 28(4), 587-604.

Ludi, S. (2002). Access for everyone: introducing accessibility issues to students in Internet programming courses. In *Frontiers in Education (FIE 2002)* (Vol. 3, pp. S1C-7-S1C-9). London, UK: IEEE Press.

Luquet, G.-H. (2001 [1927]). Children's Drawings (Le Dessin Enfantin), trans. A. Costall. London, UK: Free Association Books.

Ma, L., Ferguson, J., Roper, M., & Wood, M. (2007). Investigating the viability of mental models held by novice programmers. In *ACM SIGCSE Bulletin* 39(1), 499-503.

Mankoff, J., Fait, H., & Tran, T. (2005). Is your web page accessible?: a comparative study of methods for assessing web page accessibility for the blind. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2005)* (pp. 41-50). New York, NY, USA: ACM.

Martínez, C. C., Martínez-Normand, L., & Olsen, M. G. (2009). Is it possible to predict the manual web accessibility result using the automatic result? In Stephanidis C. (ED.) *Universal Access in Human-Computer Interaction. Applications and Services. UAHCI 2009. Lecture Notes in Computer Science, vol 5616*. Berlin/Heidelberg, Germany: Springer.

McCathie Nevile, C. (2017b). World Wide Web Consortium Process Document. Retrieved October 16, 2017 from: https://www.w3.org/2017/Process-20170301/

Meiselwitz, G., Wentz, B., & Lazar, J. (2010). *Universal Usability: Past, Present, and Future.* Hanover, MA, USA: Now Publishers Inc.

Melo, A. R., Soares, M. M., Campos, F., & Correia, W. (2013). Ergonomic evaluation of websites focusing on the human-computer interface so as to improve access to the web especially by people with visual disabilities. In Marcus A. (Ed.) *Design, User Experience, and Usability. Web, Mobile, and Product Design. DUXU 2013. Lecture Notes in Computer Science, vol 8015*. Berlin/Heidelberg, Germany: Springer.

Microsoft Corporation. (2001). Microsoft Inductive User Interface Guidelines . Retrieved October 31, 2017 from https://msdn.microsoft.com/en-us/library/ms997506.aspx

Moray, N., (1999). Mental models in theory and practice. In D. Gopher and A. Koriat (Eds.) Attention and performance XVII: cognitive regulation of performance, interaction of theory and application. Cambridge, MA, USA: MIT Press.

Moreno, L. Calvo, R., & Iglesias, A. (2011). Accessibility evaluation of Moodle centred in visual impairements. In *Proceedings of the 7th International Conference on Web Information Systems and Technologies (WEBIST 2011)* (pp.221-228). Setúbal, Portugal: Science and Technology Publications.

Moreno, L., Iglesias, A., Calvo, R., Delgado, S., & Zaragoza, L. (2012). Disability standards and guidelines for learning management systems: evaluating accessibility. In R. Babo and A. Azevedo (Eds.) *Higher Education Institutions and Learning Management Systems: Adoption and Standardization* (pp. 199-218). Hershey, PA, USA: IGI Global.

Mueller, M. J., Jolly, R., & Eggert, E. (Eds.) (2017). Policies Relating to Web Accessibility. Retrieved October 31, 2017 from https://www.w3.org/WAI/Policy/

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 15(3), 263-280.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ, USA: Prentice-Hall.

NIA & NLM. (2002). Making your Web site senior-friendly: A checklist. *National Institute on Aging and National Library of Medicine*. Retrieved October 31, 2017 from http://www.nlm.nih.gov/pubs/checklist.pdf

Nielsen, J. (2000). *Designing Web Usability*. Indianapolis, IN, USA: New Riders Publishing.

Nizar, A. A., Obedidat, A., & Abu-Addose, H. Y. (2013). Accessibility as an indicator of Jordanian E-government website quality. In Proceedings of 2013 Fourth International Conference on e-Learning (pp. 156-160). London, UK: IEEE Press.

Norman, D. A. (1983). Some observations on mental models. In D. Gentner and A. L. Stevens (Eds.), *Mental models* (p. 7-14). Hillsdale, NJ, USA: Ertbaum.

Norman, D.A., 1988. *The Psychology of Everyday Things*. New York, NY, USA: Basic Books.

Olalere, A., & Lazar, J. (2011). Accessibility of U.S. Federal Government Home Pages: Section 508 Compliance and Site Accessibility Statements. *Government Information Quarterly*, 28(3), 303-309.

Ortner, D., Batušić, M., & Miesenberger, K. (2004). Postgraduate course on accessible web design. In Miesenberger K., Klaus J., Zagler W.L., Burger D. (Eds.) *Computers Helping People with Special Needs. ICCHP 2004. Lecture Notes in Computer Science, vol 3118*. Berlin/Heidelberg, Germany: Springer.

Paciello, M. (2000). *Web Accessibility for People with Disabilities*. San Francisco, CA, USA: CMP Books.

Park, E. J., Lim, Y. W., & Lim, H. K. (2014). A study of web accessibility of websites built in HTML5 - focusing on the top 100 most visited websites. *International Journal of Multimedia and Ubiquitous Engineering*, 9(4), 247-256.

Parmanto, B., & Zeng, X. (2005). Metric for web accessibility evaluation. *Journal of the Association for Information Science and Technology*, 56(13), 1394-1404.

Payne, S.J. (1992). On the mental models and cognitive artefacts. In Rogers, Y., Rutherford, A. and Bibby, P.A. (Eds.), *Models in the Mind: Theory, Perspective and Applications* (pp.103-118). London, UK: Academic Press.

Payne, S. J. (2003). Users' mental models: The Very Ideas. In J. M. Carroll (Ed.), *HCI Models, Theories and Frameworks: Toward a multidisciplinary science*. San Francisco, CA, USA: Morgan Kaufman.

Payne, S. J. (2007). Mental models in human-computer interaction. In A. Sears and J. A. Jacko (Eds.) *The Human-Computer Interaction Handbook* (pp.63-75). Boca Raton, FL, USA: CRC Press.

Payne, S. J., Squibb, H. R., & Howes, A. (1990). The nature of device models: the yoked state space hypothesis and some experiments with text editors. *Human-Computer Interaction*, 5(4), 415-444.

Petrie, H. (Ed.) (2006). BenToWeb Project Deliverable 3.3: Report on the survey of Web site designers and commissioners of Web sites. Retrieved October 31, 2017 from http://bentoweb.org/documents.html

Petrie, H. (Ed.) (2013a). i2web Deliverable 4.7: Final user-based evaluations of accessibility and usability of services and applications.

Petrie, H. (Ed.) (2013b). i2Web Deliverable 4.8: Field studies of the use of i2Web services and applications.

Petrie, H., Hamilton, F., & King, N. (2005). Accessibility of museum, library and archive websites: the MLA audit. Centre for Human-Computer Interaction Design, City University.

Petrie, H., King, N., Velasco, C. A., Gappa, H., & Nordbrock, G. (2007). The usability of accessibility evaluation tools. In Stephanidis C. (Ed.) *Universal Access in Human-Computer Interaction. Applications and Services. UAHCI 2007. Lecture Notes in Computer Science, vol 4556* (pp.124-132). Berlin/Heidelberg, Germany: Springer.

Petrie, H., Power, C., & Swallow, D. (Eds.) (2011). i2web Deliverable 3.2: Requirements for web developers and web commissioners in ubiquitous Web 2.0 design and development. Retrieved October 31, 2017 from https://services.imergo.com/downloads/2012/201201_I2Web_D32.pdf

Petrie, H., Savva, A., & Power, C. (2015). Towards a Unified Definition of Web Accessibility. In *Proceedings of the 12th International Cross Disciplinary Conference on Web Accessibility (W4A 2015)*. New York, NY, USA: ACM.

Power, C., Freire, A., & Petrie, H. (2010). Integrating accessibility evaluation into web engineering processes. In T. Spiliotopoulous, P. Papadopoulou, D. Martakos & G. Kouroupetroglou (Eds.), *Integrating usability engineering for designing the web experience: methodologies and principles* (pp.124-148). Hershey, PA: IGI Global.

Power, C., Freire, A., Petrie, H., & Swallow, D. (2012). Guidelines are only half of the story: accessibility problems encountered by blind users on the web. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2012) (pp.433-442). New York, NY, USA: ACM.

Power, C., Petrie, H., Freire, A. P., & Swallow, D. (2011). Remote evaluation of WCAG 2.0 techniques by web users with visual disabilities. In Stephanidis C. (Ed.) *Universal Access in Human-Computer Interaction. Design for All and eInclusion. UAHCI 2011. Lecture Notes in Computer Science, vol 6765*. Berlin/Heidelberg, Germany: Springer.

Putnam, C., Wozniak, K., Zefeldt, M. J., Cheng, J., Caputo, M., & Duffield, C. (2012). How do professionals who create computing technologies consider accessibility?. In Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2012) (pp.87-94). New York, NY, USA: ACM.

Quin, L. (Ed.) (2013). HTML Accessibility Task Force Work Statement. Retrieved October 31, 2017 from https://www.w3.org/WAI/PF/html-task-force

Raggett, D., Le Hors, A., & Jacobs, I. (Eds.) (1997). HTML 4.0 Specification – W3C Recommendation 18-Dec-1997. Retrieved October 31, 2017 from https://www.w3.org/TR/REC-html40-971218/

Richards, J., Spellman, J., & Treviranus, J. (Eds.) (2015). Authoring Tool Accessibility Guidelines (ATAG) 2.0 – W3C Recommendation 24 September 2015. Retrieved October 31, 2017 from https://www.w3.org/TR/ATAG20/

Richardson, M., & Ball, L. J. (2009). Internal representations, external representations and ergonomics: towards a theoretical integration. *Theoretical Issues in Ergonomics Science*, 10(4), 335-376.

Rips, L. J. (1986). Mental Muddles. In M. Brand, & R. M. Harnish (Eds.) *The Representation of Knowledge and Belief* (pp. 258-286). Tuscon, AZ, USA: University of Arizona Press.

Rogers, M. (2017). Government accessibility standards and WCAG 2. Retrieved October 31, 2017 from http://www.powermapper.com/blog/government-accessibility-standards/

Rømen, D., & Svanæs, D. (2008). Evaluating web site accessibility: validating the WAI guidelines through usability testing with disabled users. In *Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges (NordiCHI 2008)* (pp.535-538). New York, NY, USA: ACM.

Rømen, D., & Svanæs, D. (2012). Validating WCAG versions 1.0 and 2.0 through usability testing with disabled users. *Universal Access in the Information Society*, 11(4), 375-385.

Rosencrance, L. (2001). W3C recommends online accessibility guidelines. *CNN.com*. Retrieved October 31, 2017 from http://edition.cnn.com/2001/TECH/internet/09/20/internet.disabilities.idg/index.html

Rosmaita, B. J. (2006). Accessibility first!: a new approach to web design. *ACM SIGCSE Bulletin*, 38(1), 270-274.

Rosson, M. B., Ballin, J. F., Rode, J., & Toward, B. (2005). "Designing for the Web" revisited: a survey of informal and experienced web developers. In D. Lowe and M. Gaedke (Eds.), *Web Engineering. ICWE 2005. Lecture Notes in Computer Science, vol 3579* (pp.522-532). Berlin/Heidelberg, Germany: Springer.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, 100(3), 349.

Sloan D., Gregor P., Rowan M., & Booth P. (2000). Accessible accessibility. In Scholtz J. and Thomas J. (Eds.). Proceedings of the *2000 Conference on Universal Usability (CUU 2000)* (pp.96-101). New York, NY, USA: ACM.

Sloan D., Kelly B., Heath A., Petrie H., Hamilton F., & Phipps L. (2006). Contextual Accessibility: Maximizing the Benefit of Accessibility Guidelines. In *Proceedings of the 2006 International Cross Disciplinary Workshop on Web Accessibility (W4A 2006)* (pp.121-131). New York, NY, USA: ACM.

Sloan, D., & Kelly, B. (2008). Reflections on the development of a holistic approach to web accessibility. In H. Petrie (Ed.) Proceedings of Accessible Design in the Digital World (ADDW 2008).

Sloan, D., Gregor, P., Booth, P., & Gibson, L. (2002). Auditing accessibility of UK Higher Education web sites. *Interacting with Computers*, 14(4), 313-325.

Stewart, R., Narendra, V., & Schmetzke, A. (2005). Accessibility and usability of online library databases. *Library Hi Tech*, 23(2), 265-286.

Sullivan, T., & Matson, R. (2000). Barriers to Use: Usability and Content Accessibility on the Web's Most Popular Sites. In Scholtz J. and Thomas J. (Eds.). Proceedings of the *2000 Conference on Universal Usability (CUU 2000)* (pp.139-144). New York, NY, USA: ACM.

Tatomir, J., & Durrance, J. C. (2010). Overcoming the information gap: Measuring the accessibility of library databases to adaptive technology users. *Library Hi Tech*, 28(4), 577-594.

Thatcher, A., & Greyling, M. (1998). Mental models of the Internet. *International Journal of Industrial Ergonomics*, 22(4), 299-305.

Thatcher, J., Burks, M. R., Heilmann, C., Henry, S. L., Kirkpatrick, A., Lauke, P. H., Lawson, B., Regan, B., Rutter, R., Urban, M., & Waddel, C. D. (2006). *Web Accessibility: Web Standards and Regulatory Compliance*. Berkeley, CA, USA: Apress.

Thompson, T., Burgstahler, S., & Moore, E. J. (2007). Accessibility of Higher Education Web pages in the Northwestern US: Current Status and Response to Third Party Outreach. In Proceedings of the First International Conference on Technology-based Learning with Disability (p. 127).

Thompson, T., Burgstahler, S., & Moore, E. J. (2010). Web accessibility: A longitudinal study of college and university home pages in the northwestern United States. *Disability and Rehabilitation: Assistive Technology*, 5(2), 108-114.

Thompson, T., Burgstahler, S., Moore, E., Gunderson, J., & Hoyt, N. (2007). International research on web accessibility for persons with disabilities. In M. Khosrow-Pour (Ed.), *Managing Worldwide Operations and Communications with Information Technology*. Hershey, PA: Information Resources Management Association.

Timmins, F., & McCabe, C. (2005). How to conduct an effective literature search. *Nursing Standard*, 20(11), 41-47.

Treviranus, J., McCathieNevile, C., Jacobs, I., & Richards, J. (Eds.) (2000). Authoring Tool Accessibility Guidelines 1.0 – W3C Recommendation 3 February 2000. Retrieved October 31, 2017 from https://www.w3.org/TR/WAI-AUTOOLS/

Trewin, S., Cragun, B., Swart, C., Brezin, J., & 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 2010). New York, NY, USA: ACM Press.

Turner, P., & Sobolewska, E. (2009). Mental Models, Magical Thinking and. Individual Differences. *Human Technology*, 5(1), 90-113.

UK Cabinet Office. (2005). eAccessibility of public sector services in the European Union. London: Cabinet Office. Archived from the original on 28 April 2006. Retrieved 31 October, 2017 from http://webarchive.nationalarchives.gov.uk/20060715170044/http://www.cabinetoffice.gov.uk/e-government/resources/eaccessibility/

UK Government (2010). Equality Act 2010. London: HMSO. Retrieved October 31, 2017 from https://www.legislation.gov.uk/ukpga/2010/15/contents

United Nations (2006). Convention on the Rights of Persons with Disabilities. Retrieved October 31, 2017, from http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/61/106

United States Access Board (2016). Access Board Approves Rules on ICT Refresh and Medical Diagnostic Equipment. Retrieved 31 October, 2017 from https://www.access-board.gov/news/1824-access-board-approves-rules-on-ict-refresh-and-medical-diagnostic-equipment

US Government (1998). Section 508 of the Rehabilitation Act (29 U.S.C. 794d), as amended by the Workforce Investment Act of 1998 (P.L. 105-220). Retrieved October 31, 2017 from http://www.section508.gov

US Government. (1990). Americans with Disabilities Act of 1990 (P.L. 101-336). Washington, DC, USA: Government Printing Office.

Vanderheiden, G. C. (1995). Design of HTML (Mosaic) Pages to Increase their Accessibility to Users with Disabilities Strategies for Today and Tomorrow. Retrieved

October 31, 2017 from https://trace.umd.edu/publications/design-html-mosaic-pages-increase-their-accessibility-users-disabilities-strategies

Vanderheiden, G. C., & Chisholm, W. A. (1996). Design of HTML Pages to Increase their Accessibility to Users with Disabilities Strategies for Today and Tomorrow. Retrieved October 31, 2017 from http://trace.wisc.edu/archive/html_guidelines/version6.8.html

Vanderheiden, G. C., & Chisholm, W. A. (1998). Unified Web Site Accessibility Guidelines 8.0. Retrieved October 31, 2017 from www.w3.org/WAI/GL/central.htm

Velleman, E., & Abou-Zahra, S. (Eds.) (2014). Website Accessibility Conformance Evaluation Methodology (WCAG-EM) 1.0 – W3C Working Group Note 10 July 2014. Retrieved October 31, 2017 from https://www.w3.org/TR/WCAG-EM/

Velleman, E., Velasco, C., Snaprud, M., & Burger, D. (Eds.) (2006). Unified Web Evaluation Methodology (UWEM 1.0). Retrieved October 31 2017 from http://www.wabcluster.org/uwem1/UWEM_1_0.pdf

Vigo, M., & Brajnik, G. (2011). Automatic web accessibility metrics: Where we are and where we can go. *Interacting with Computers*, 23(2), 137-155.

Vigo, M., Arrue, M., Brajnik, G., Lomuscio, R., & Abascal, J. (2007). Quantitative metrics for measuring web accessibility. In *Proceedings of the 2007 International Cross Disciplinary Conference on Web Accessibility (W4A 2007)* (pp.99-107). New York, NY, USA: ACM.

Vigo, M., Brown, J., & Conway, V. (2013). Benchmarking web accessibility evaluation tools: measuring the harm of sole reliance on automated tests. In *Proceedings of the 10th International Cross Disciplinary Conference on Web Accessibility (W4A 2013)* (p. 1). New York, NY, USA: ACM.

Vora, P.R. (1998). Design/Methods & Tools: Designing for the Web: a survey. *ACM Interactions*, 5(3), 13-30.

W3C/WCAG Working Group (2016). Designing Silver. (Wiki page). Retrieved October 31, 2017 from https://www.w3.org/WAI/GL/wiki/Designing_Silver

WaSP InterAct. (n.d.). About. Retrieved October 31, 2017 from http://teach.webstandards.org/about/

"WCAG 2.0 Conformance" (n.d.). Retrieved October 31, 2017 from https://www.section508.gov/content/build/website-accessibility-improvement/WCAG-conformance

"Web accessibility". (n.d.). In *Wikipedia, The Free Encyclopedia*. Retrieved October 31, 2017, from https://en.wikipedia.org/wiki/Web_accessibility

Web Standards Project. (n.d.). ATF Manifesto. Retrieved October 31, 2017 from http://www.webstandards.org/action/atf/manifesto/

WebAIM. (2014). Introduction to Web Accessibility. Retrieved October 31, 2017 from http://webaim.org/intro/

Welford, A.T. (1961). On the human demands of automation: mental work conceptual model, satisfaction and training. *Industrial and Business Psychology*, 5, 182–193.

Wilson, J. R., & Rutherford, A. (1989). Mental models: Theory and application in human factors. *Human Factors*, 31(6), 617-634.

Yesilada, Y., Brajnik, G., Vigo, M., & Harper, S. (2015). Exploring perceptions of web accessibility: a survey approach. *Behaviour & Information Technology*, 34(2), 119-134.

Young, R. M. (1981). The machine inside the machine: users' models of pocket calculators. *International Journal of Man-Machine Studies*, 15(1), 51-85.

Young, R.M. (1983). Surrogates and mappings: two kinds of conceptual models of interactive devices. In D. Gentner and A.L. Stevens, (Eds.). *Mental models*. Hillsdale, NJ, USA: Ertbaum.

Yu, D. X., & Parmanto, B. (2011). US state government websites demonstrate better in terms of accessibility compared to federal government and commercial websites. *Government Information Quarterly*, 28(4), 484-490.

Yu, H. (2002). Web accessibility and the law: recommendations for implementation. *Library Hi Tech*, 20(4), 406-419.

Appendix A. Informed consent form

This research study is investigating *[the effectiveness of different accessibility information resources]*. The aim of the study is to *[determine how effective each accessibility information resource is in helping web developers identify and resolve accessibility problems]*.

The entire session will last approximately *[one hour]*. You will receive *[a £20 Amazon gift voucher]* for taking part in the study.

With your consent, the session will be *[audio/video]* recorded. The *[audio/video]* recordings will be transcribed by David Swallow. You will be assigned an anonymous ID (e.g.”WD37”) to identify your responses within the research study. You will not be identified when this data is shared, described or interpreted.

All data gathered will be treated confidentially. Your data will be stored securely on a password protected computer. Only David Swallow and his PhD supervisors, Dr. Christopher Power and Prof. Helen Petrie, will have access to the data collected today in its original format.

Please check the boxes if you agree with the following statements:

- ☐ I confirm that I have read and understand the pre-study information and have had the opportunity to ask any questions.
- ☐ I understand that my participation in this study is voluntary and that I am free to withdraw at any time.
- ☐ I understand that data collected about me during this study will be anonymised and I will not be identified when it is shared, described or interpreted.
- ☐ I agree to allow the data collected to be used in David Swallow's PhD thesis and possibly other research publications.
- ☐ I agree to take part in this study.

Please *[sign your name/enter your email address]*, below, to indicate your consent to participate in the study.

______________________	_____/_____/_____
Signature of research participant	Date

Appendix B. Contextual inquiry schedule

The contextual inquiry schedule covered the following topics:

- The **web developer:** their knowledge and skills, their organisation, and the nature of the work they are involved in.
- **Workspace configuration:** an "introduction" to the participant's workspace and immediate surroundings.
- **Communication:** who the web developer communicates with during a typical day, and how.
- **Help and information:** who and what resources the web developer turns to when they need help.
- **Standards-compliance:** what standards-compliance means to the web developer and how they achieve it.
- **Preview:** how the web developer previews a website they have created.
- **Validation:** how the web developer validates a website they have created.
- **Users:** what consideration the web developer gives to the users of their websites.
- **Accessibility:** what consideration the web developer gives to the accessibility of their websites.
- **Accessibility testing:** how the web developer tests the accessibility of a website they have created.
- **Future improvements:** how the web developer could improve the accessibility of the website they have created in future.

Appendix C. Mental models interview schedule

The following questions aim to establish whether the developer currently attempts to make their websites accessible, whether they conduct any accessibility testing and how they go about it, and whether their clients have any accessibility requirements or expectations.

- Do you attempt to make your websites accessible?
- What sort of things do you do to make your websites accessible?
- What factors motivate you to take accessibility into account?
- What factors prevent you from taking accessibility into account?
- What resources do you use when making your website accessible?
- Are you familiar with WCAG (Web Content Accessibility Guidelines)?
- Have you ever tried to use WCAG?
- How easy or difficult do you find WCAG to follow?
- Are you familiar with the three conformance levels of WCAG?
- Do you ever test to any of the WCAG conformance levels?
- Why and in what circumstances do you do conformance testing?
- Do you test your websites with disabled users and/or older users?
- Why and in what circumstances do you do user testing?
- How many users do you test with and from what groups?

The following questions refer to the people who commission websites. These may be external clients, people in their organisation, or the developer's manager):

- Do the people who commission your websites require/expect you to make your websites accessible?
- Do the people who commission your websites require/expect you to meet a particular standard of accessibility (e.g. WCAG Level A)?
- Do the people who commission your websites require/expect you to test the accessibility of your websites?
- Do the people who commission your websites test the accessibility of your websites themselves?

Appendix D. Initial WebAIR evaluation interview schedule

- What are your immediate impressions of the resource?
- How useful did you find the resource?
- How easy or difficult did you find the resource?
- How understandable did you find the resource?
- How clear was the layout of the individual pages in the resource?
- How easy or difficult was it to navigate the resource?
- Do you consider it a good idea to organise the resource in this way?
- Does the organisation of the resource make sense to you?
- Do the organisation of the resource relate to the way in which you work?
- Are the number of categories too much, too little, or about right?
- Are there any categories that you feel are missing?
- Are there any categories that you feel are unnecessary or irrelevant?
- Was the amount of information in the resource too much, too little, or just right?
- Is there any information that you feel is missing?
- Is there any information that you feel is unnecessary or irrelevant?
- Was the number of things to do in the resource too much too little or about right?
- Are there any things to do that you feel are missing?
- Is there any things to do that you feel are unnecessary or irrelevant?
- To what extent has your understanding of accessibility been increased by the resource?
- How confident would you be about creating accessible websites using the resource?
- How likely are you to use the resource?
- Under what circumstances might you use the resource?

- Having used each of the different resources, what are your overall impressions?
- Which of the four resources are you likely to use in future, and why?

Appendix E. "In the wild" evaluation diary entry survey

Please complete the following information to provide feedback about your experience of WebAIR. You may complete this as many times as you wish during the two-week study period.

- Briefly, what were you working on at the time you used WebAIR? (open-ended)
- For what purpose did you use WebAIR?
    - To look up a specific accessibility issue (e.g. alt descriptions on images)
    - To evaluate a particular content type (e.g. a form)
    - To evaluate an entire webpage/website
    - For general information/educational purposes
    - Other (please specify)
- What sections of WebAIR did you look at?
    - Forms
    - Links
    - Tables
    - Images
    - Text
    - Audio & Video
    - Time Limits
    - Navigation
- Were you able to find what you were looking for?
    - Yes
    - No
- Do you have any comments or suggestions for how we could improve WebAIR? (open-ended)

Appendix F. "In the wild" evaluation final survey

Thank you for taking part in the WebAIR evaluation study. You have been using WebAIR for approximately two weeks now, in the creation of an accessible webpage. Please complete the following questions about WebAIR.

- How confident are you about the accessibility of the webpage you have created using WebAIR? - Not at all confident [1 2 3 4 5] Very confident
- Approximately how long did it take you to create an accessible webpage using WebAIR? (open-ended)
- To what extent, if at all, did making the webpage accessible using WebAIR add to the time it would normally take? - Not at all [1 2 3 4 5] Very much
- What characteristics of WebAIR (e.g. its formatting, layout, design, organisation etc.) do you feel either helped or hindered you? (Please explain)
- What types of accessible web content (e.g. tables, images, forms etc.) do you feel either sped you up or slowed you down? (Please explain)

Please rate WebAIR on the following attributes:

- Usefulness - Very low [1 2 3 4 5] Very high
- Ease of use - Very difficult [1 2 3 4 5] Very easy
- Navigable - Very low [1 2 3 4 5] Very high
- Understandable - Very low [1 2 3 4 5] Very high
- Completeness - Very low [1 2 3 4 5] Very high
- Amount of information provided - Far too little [1 2 3 4 5] Far too much
- Number of items to test (e.g. the number of things that WebAIR requires you to do) - Far too few [1 2 3 4 5] Far too many
- Organisation - Very unclear [1 2 3 4 5] Very clear
- Likelihood of using - Very unlikely [1 2 3 4 5] Very likely

- Was there any information that you feel is missing from WebAIR? (open-ended)
- Was there any information in WebAIR that you feel is unnecessary or irrelevant? (open-ended)
- Was there any information in WebAIR that you feel is incorrect? (open-ended)

- To what extent, if at all, has your understanding of web accessibility been increased by WebAIR? - Not at all [1 2 3 4 5] Very much
- Under what circumstances might you use WebAIR? (If you would not use WebAIR, please explain why)
- In addition to the comments that you (may) have provided in your individual diary entries, do you have any comments about WebAIR or about the study in general? (open-ended)

Thank you once again for your participation in this study.

Appendix G. Accompanying material

The following material can be found on the CD-ROM accompanying this thesis:

A. Systematic review references
B. Contextual inquiry verbatim transcripts
C. Contextual inquiry verification questions
D. Mental models interview verbatim transcripts
E. Mental models interview verification codes
F. WebAIR to WCAG 2.0 mappings
G. Professional web developer evaluation interview verbatim transcripts
H. Professional web developer evaluation verification codes
I. Novice web developer evaluation training document
J. Novice web developer evaluation protocol
K. Novice web developer evaluation questionnaire
L. Student web developer evaluation materials
M. Professional web developer evaluation sample notes
N. Professional web developer evaluation verification codes
O. "In the wild" evaluation website specification
P. "In the wild" evaluation participant webpages