

# Guidelines are Only Half of the Story: Accessibility Problems Encountered by Blind Users on the Web

Christopher Power, André Pimenta Freire, Helen Petrie, David Swallow

Department of Computer Science, University of York

Deramore Lane, York, YO10 5GH, UK

{cpower, apfreire, helen.petrie, dswallow}@cs.york.ac.uk

## ABSTRACT

This paper describes an empirical study of the problems encountered by 32 blind users on the Web. Task-based user evaluations were undertaken on 16 websites, yielding 1383 instances of user problems. The results showed that only 50.4% of the problems encountered by users were covered by Success Criteria in the Web Content Accessibility Guidelines 2.0 (WCAG 2.0). For user problems that were covered by WCAG 2.0, 16.7% of websites implemented techniques recommended in WCAG 2.0 but the techniques did not solve the problems. These results show that few developers are implementing the current version of WCAG, and even when the guidelines are implemented on websites there is little indication that people with disabilities will encounter fewer problems. The paper closes by discussing the implications of this study for future research and practice. In particular, it discusses the need to move away from a problem-based approach towards a design principle approach for web accessibility.

## Author Keywords

Web accessibility; web accessibility guidelines; user evaluation; blind users.

## ACM Classification Keywords

H.5.2 [User interfaces]: Evaluation/methodology;  
H.5.4 [Hypertext/Hypermedia]: User issues;

## INTRODUCTION

In the information society, the Web provides people with the ability to tap into news, commerce and social information at any time. Indeed, one could say that users are drowning in information. However, not all people can use this vast resource of information equally. The persistence of websites that are not accessible, meaning that people with disabilities cannot use them [15], results in these users are living in an information desert in comparison to their mainstream peers [22]. Even worse, recent studies show that the Web is becoming less accessible to people with disabilities over time [16].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI'12, May 5–10, 2012, Austin, Texas, USA.

Copyright 2012 ACM 978-1-4503-1015-4/12/05...\$10.00.

In 1999 the Web Accessibility Initiative (WAI) of the World Wide Web Consortium (W3C) defined the first version of the Web Content Accessibility Guidelines (WCAG 1.0). WCAG 1.0 was published to promote web accessibility and to provide a comprehensive set of guidelines on how to prepare web content so that people with disabilities could use the web regardless of their needs and preferences [32]. WCAG 1.0 comprised 14 guidelines, and, within these guidelines, 65 checkpoints (CPs) that described how developers could adapt their web content in order to make it accessible. Each checkpoint was assigned a priority level, Priority 1 through Priority 3, which indicated the importance of the CP in terms of its impact on accessibility of content to different groups of disabled users [31]. If a web page satisfied all Priority 1 CPs it was said to be conformant at Level A. Likewise, if a website satisfied all Priority 1 and 2 CPs, the website was conformant at Level AA. Finally, if a website satisfied all CPs, it was conformant to Level AAA.

For almost a decade WCAG 1.0 served as the *de facto* standard for web accessibility. It can be argued that WAI's original goal of raising awareness accessibility was achieved [17], with WCAG 1.0 becoming the basis for legislation in a number of countries [2], and it is heavily referenced in web accessibility practice [28, 31].

However, the impact of WCAG 1.0 on improving the accessibility of the Web remained quite low throughout the period of its use. Evaluations using automated tools covering a small subset of the guidelines [13, 14, 33], expert evaluations using a combination of automated testing tools and human judgment [19, 30], and user evaluations with disabled participants [4, 10, 24] all found the level of accessibility of web pages to be extremely low in both the public and private sectors.

This low level of accessibility is likely to be the result of several different factors. For instance, despite awareness of accessibility increasing over the last decade at the level of government and legislation, the level of knowledge in the community of web commissioners and web masters remains quite low. In 2005, Petrie et al. [24] reported that 30% of websites claiming conformance to some level of WCAG 1.0 overstated their level of conformance. The authors hypothesized that this may be that web

commissioners do not understand the differences between automated and manual testing, and therefore attribute more importance to the former when making conformance claims. This hypothesis is supported by the work by Lazar et al. [18], who found that the role of accessibility tools and related guidelines remained unclear to many of the site owners they surveyed. The most alarming aspect of the Lazar et al. study was the fact that 22% of site owners had no knowledge at all about accessibility guidelines.

In addition to this, there have been several criticisms leveled at the provision of support to developers for testing websites for accessibility based on WCAG 1.0. WCAG 1.0 was evaluated with 12 experienced web developers who were novices to accessibility [7]. This evaluation demonstrated that developers struggled with both the navigation through WCAG documents as well as the language used in them. Other work has also pointed to ambiguities in the language, as well as the required level of technical knowledge of accessibility in order to read and interpret the guidelines, as being barriers to their uptake in accessibility practice [11].

However, the most obvious and important concern about WCAG 1.0 was the lack of empirical evidence that a website which is Level AAA conformant was more usable by people with disabilities than a Level A website. There are two examples of empirical data being collected regarding WCAG 1.0. Rømen and Svanæs [29] collected data with 7 users on 2 websites and found that only 27% of problems identified by users were covered by the guidelines. Secondly, the Disability Rights Commission (DRC) [10] conducted a Formal Investigation into access and inclusion on the Web. The result of this investigation is still the largest known accessibility evaluation to date with 1000 websites being evaluated with automated testing and 100 websites being evaluated by experts and disabled users. For the user evaluations, 913 tasks were undertaken over the 100 websites chosen, with participants being selected from a wide variety of people with disabilities. Only 19% of websites met even Level A conformance to WCAG 1.0. In the user evaluations, it found that of the problems reported by users 45% of them were not covered by WCAG 1.0 CPs. As a consequence of these results, the equating of accessibility to conformance to WCAG 1.0 is highly suspect, and the accessibility community has indicated that this is an important issue to be addressed [17].

In 2008 the WAI released a new version of WCAG. WCAG 2.0 attempted to address many of the criticisms that had been leveled at WCAG 1.0. These new guidelines [6] are organized in a more hierarchical manner. Within WCAG 2.0 there are four Principles about web content accessibility: content should be Perceivable, Operable, Understandable and Robust. These principles group guidelines in a more structured hierarchy than was present in WCAG 1.0. Further, Guidelines under each of these Principles have been rephrased to be solutions to specific

user requirements, such as the provision of text alternatives for non-text content (Guideline 1.1). Then, for each Guideline there are Success Criteria (SC). SCs are testable statements that a developer can use to determine if web content is accessible. It is against these SC that a website is measured for conformance, with each SC having a priority level, Level A, AA or AAA, relating to conformance levels that are similar to WCAG 1.0.

In order to future-proof WCAG 2.0 against the fast evolution of technology, the WAI removed the technical aspects of accessibility from the Guidelines and SCs. Technical information regarding how to implement web content with existing web technologies is now provided in separate documents [8]. These documents describe techniques that have been determined by the WCAG Working Group to be *sufficient* to meet the SC. It is also possible to meet an SC without using these sufficient techniques. For example, developers can provide implementations and evaluate them with users to show they pass the SC. However, it is reasonable to expect that the majority of web developers will implement the sufficient techniques in their own websites. In addition to these organizational changes, the WAI has provided multiple documents for understanding accessibility, WCAG and other related concepts (e.g. user agents, authoring tools) [9].

However, despite all of these changes, studies have demonstrated that many of the problems from WCAG 1.0 persist in WCAG 2.0. Petrie et al. [26] conducted interviews with 14 web accessibility evaluators, and found that they were very unclear on the differences between automated testing and manual testing of accessibility, in particular on what can be tested through automated tools. Brajnik et al. [5] had 22 expert and 27 non-expert evaluators perform accessibility evaluations using WCAG 2.0 and found that for 50% of SCs evaluators were unable to come to an 80% level of agreement about whether a problem was present in a page. Further, when using WCAG 2.0 20% of the problems reported by expert evaluators were false positives and they also missed 32% of true accessibility problems. For non-experts, the results were even worse, with higher levels of both false positives and false negatives. Alonso et al. [3] also found that a group of 25 beginner evaluators struggled to consistently rate problems according to WCAG 2.0. One cause of this lack of success in identifying accessibility problems was in the interpretation of the Guidelines and SCs. All of these results demonstrate that there are understandability and interpretation problems with WCAG 2.0.

Of greater concern is that in the three years following the release of WCAG 2.0, there appears to have been little improvement in the level of web accessibility. Lopes et al. [22] crawled 30 million web pages, performing automated accessibility tests. Under 4% of elements on these web pages met all relevant SCs that could be tested automatically. This is not a very good result given that

automated testing covers only a small proportion of WCAG 2.0.

Beyond all of these issues, there is still the looming problem of a lack of empirical evidence that demonstrates that conformance to WCAG 2.0 leads to more accessible websites for disabled users. This is in contrast to web usability guidelines such as the widely used guidelines from the U.S. Department of Health and Human Sciences that cite empirical research to establish the strength of evidence supporting their application [1], and to accessibility guidelines that have been defined and validated for specific user groups [20, 21]. Indeed, there are very few examples of empirical studies of WCAG 2.0. To date, only one known paper, that of the current authors, provides (remote) user evaluations of one of the WCAG 2.0 SCs and related techniques [27].

This paper presents a study that addresses these issues and answers the following empirical questions on the relationship between WCAG conformance and accessibility as actually experienced by disabled users:

- For the web accessibility problems encountered by people with visual disabilities, what percentage of those problems would be addressed by correctly implementing existing SCs in WCAG 2.0?
- Do blind users encounter accessibility problems on web pages when they have conformed to WCAG 2.0 SCs?

## METHOD

### Design

The study involved a task-based user evaluation with blind users on selected websites and collecting information about the problems they encounter. Participants undertook a concurrent think aloud protocol, where they spoke about what they were thinking and doing as they were carrying out their tasks. Participants were asked to rate the severity of each problem they encountered using a four point scale (Cosmetic, Minor, Major, Catastrophic), adapted from the severity rating scale defined by Nielsen for usability problems [23]. A total of 10 different users, assigned from the participant group in no particular order, evaluated each of 16 websites.

### Participants

Thirty-two blind participants took part, of whom 22 were male and 10 were female. Their ages ranged from 18 to 65 years (median = 39 years). 17 participants had no residual vision, 12 had only light/dark perception and three had a very little central vision. Most participants (20 out of 32) had been blind since birth, and the remainder of the participants had been blind for a period between three years and 47 years.

All participants used screenreaders as their primary assistive technology to access computers, 30 out of 32 used JAWS and 2 used WindowEyes. The WindowEyes users

used version 7.11. JAWS versions varied from as early as JAWS 5.0 to JAWS 11.0 (the most up-to-date version available when the study was conducted). 12 users rated themselves experts with their screenreader, 6 as advanced users, 12 as intermediate and 2 as having only a basic command of their screenreader.

The participants rated their experience with computers on a scale from 1 (not at all) to 7 (extensive). Their ratings of computer experience ranged from 4 to 7, with 87% of the participants rating their experience as 5 or above. Most of the participants (29 out of 32) had been using the Internet for seven years or more. Internet Explorer was the most popular browser used by participants, being mentioned as primary navigator by all but one participant, who used Firefox.

Each participant was reimbursed with £15 per hour for his/her participation in the study.

### Equipment and software

The evaluations were performed using a personal computer running Windows XP Operating System (Service Pack 3), equipped with speakers, keyboard and 2-button mouse with scrollwheel. Users had the choice of either version 10.0 of the JAWS screenreader or version 7.11 of the WindowEyes screenreader. Participants also could choose one of two web browsers: Internet Explorer 8.0 or Firefox 3.5. Morae 3.1 was used to record the participants with their concurrent verbal protocol, their facial expressions and the computer desktop.

### Websites and tasks

Seventy-two websites from the DRC study [10] that still existed were chosen as the initial pool of websites to be used in this study. The home page of each website was manually audited to establish its conformance to WCAG 1.0 and WCAG 2.0. Only the home page was audited as the DRC study found that the conformance of the home page correlates highly with other pages on a website. However, very few of these websites met WCAG 1.0 or 2.0 conformance on initial testing, so further websites that claimed good conformance to WCAG 1.0 or 2.0 were sought from many sources.

A set of 16 websites was selected for inclusion in the study using the conformance results. This set of websites included websites from the private and public sectors, local and central government, public services, non-profit organizations and commercial websites. Unfortunately, in spite of seeking widely for websites with conformance to WCAG 1.0 Level AAA, WCAG 2.0 Level AA and Level AAA, very few such sites could be found. Table 1 shows the list of websites and information about their level of conformance to WCAG 1.0 and WCAG 2.0. For each website, the level of conformance (column *Conf.* with values A, AA, AAA and Fail for failure) is shown, along with the number of CPs/SCs violated and the number of *instances* of violations (column *Inst.*)

The *instances* of violations of CPs/SCs (*Inst.*) is the number of times each CPs/SCs is violated. For example, a website has two different CPs/SCs violated: there are 10 images without alternative text (CP 1.1/SC 1.1.1) and 22 missing headings (CP 5.1/SC 2.4.6). The *instances* of violations is 32.

Table 1 shows that only 1 website with WCAG 2.0 Level AAA and 1 website with WCAG 2.0 Level AA could be found. In addition, a number of sites, which conformed to WCAG 1.0 at A or AA, failed WCAG 2.0 conformance. When examining the ways in which websites failed to conform, it was noted that 3 of the websites (Digizen, JISC and Royal Mail) failed one SC a single time, specifically SC 3.3.1 (“error identification”), and otherwise conformed to Level A of WCAG 2.0. Therefore, in some analyses, which will be noted, these websites are classified as Level A conformant websites.

For each website, two or three typical tasks of varying levels of difficulty were created. Tasks involved simple operations such as finding information about council tax, booking tickets for a concert and finding an exhibit on display in a large museum. Tasks required between two and 8 steps to be completed, with an average of four steps per task.

**Procedure**

Evaluations took place in the Interaction Laboratory in the Department of Computer Science at the University of York. Participants were briefed about the study and asked to sign an informed consent form. Participants, with assistance from the researcher when needed, adjusted the browser and screenreader to the participant’s preferred settings. Each website was introduced to the participants and they undertook tasks on the website while providing a concurrent verbal protocol. When participants were quiet for an extended period of time, they were prompted for their thoughts by the researcher. When participants encountered a problem, the researcher asked them to pause briefly and rate the problem for severity using the four point scale. This procedure was repeated for each website. At the end of the session, participants were debriefed, and were invited to ask questions about the study. All web pages visited by users were archived for further analysis.

**Data analysis**

Three coders independently identified and categorized accessibility problems on a subset of videos from the evaluation sessions.

Website	WCAG 1.0			WCAG 2.0			User Problems				
	Conf.	CPs violated	Inst.	Conf.	SCs violated	Inst.	Cos.	Min.	Maj.	Cat.	Total
www.lflegal.com	AA	2	5	AAA	0	0	6	27	12	6	51
www.green-beast.com	AA	3	23	AA	3	9	11	24	11	2	49
www.york.gov.uk	AA	4	16	Fail	5	7	5	29	25	11	70
www.nhsns.org	AA	6	30	Fail	9	31	10	29	32	21	92
www.copac.ac.uk	A	8	21	A	2	6	9	25	17	6	57
www.theaa.com	A	9	68	Fail	9	58	7	22	20	5	54
www.dh.gov.uk	A	9	91	A	6	31	6	44	30	14	94
www.digizen.org.uk	A	9	80	Fail	12	46	2	6	15	31	54
www.jisc.ac.uk	A	12	58	Fail	13	216	5	12	19	15	51
www.royalmail.com	A	15	50	Fail	7	103	5	23	19	7	54
www.pret.co.uk	A	16	184	Fail	21	141	13	24	14	3	54
www.tuc.org.uk	A	23	146	Fail	17	97	17	36	34	13	100
ww.britishmuseum.org	Fail	8	130	Fail	8	86	16	73	44	23	156
www.nhsdirect.nhs.uk	Fail	10	30	Fail	20	163	14	36	21	9	80
www.ford.co.uk	Fail	27	124	Fail	33	244	19	44	57	49	169
www.ticketmaster.co.uk	Fail	29	757	Fail	35	1118	12	96	62	28	198

**Table 1. List of websites evaluated in the study, with conformance levels to WCAG 1.0 and WCAG 2.0, number of different CPs/SCs violated and instances of violations, and instances of user problems grouped by severity levels**

The three coders compared their initial set of user problems and their categorizations in order to develop a unified list of user problems and user problem categories. Several iterations of coding and discussion were needed before a final set of categories was agreed. Then, using this set of categories, the main coding of all the user sessions was performed.

For each user problem, an analysis was conducted as to whether it had one or more *relevant* WCAG 1.0 CPs and/or WCAG 2.0 SCs (CPs/SCs). A set of guidelines *covered* a user problem if one or more CPs/SCs was identified to be *directly relevant* to the user problem, meaning it was clear that it addressed the problem encountered by the user. In the analysis, some user problems were identified as having *marginally relevant* CPs/SCs, meaning that the CPs/SCs could be interpreted as addressing the user problem from a certain point of view, but it was not totally clear that it was relevant. For each user problem, the web page on which it was encountered was evaluated to see if it passed the *directly relevant* and *marginally relevant* CPs/SCs.

## RESULTS

Table 1 shows the number of user problems for each website, identified at each severity level: cosmetic (*Cos.*), minor (*Min.*), major (*Maj.*) and catastrophe (*Cat.*). The participants encountered 1383 instances of accessibility problems across the 16 websites. The mean number of problems per website was 86.4.

### User problems and WCAG conformance

Analyses were performed on how the number of problems encountered by users relate to the conformance of the websites to both WCAG 1.0 and WCAG 2.0.

For WCAG 1.0, Figure 1 presents the mean number of problems on websites that were non-conformant, Level A or Level AA conformant (there being no Level AAA conformant websites for WCAG 1.0). There was a significant difference between the mean number of problems found on websites with the different conformance levels ( $F = 12.35$ ,  $df = 2, 13$ ,  $p < 0.001$ ). A set of Tukey HSD post-hoc tests showed that the difference between non-conformant and Level A conformant websites was significant ( $p < 0.001$ ) and that the difference between non-conformant and Level AA websites was also significant ( $p < 0.005$ ).

For WCAG 2.0, Figure 2 shows the mean number of problems on websites that were non-conformant, conformant to Level A and conformant at any level. An interesting observation is that three of the non-conformant websites that failed Level A by only violating SC 3.3.1 had some of the fewest user problems (see Table 1 entries Digizen - [www.digizen.org.uk](http://www.digizen.org.uk), JISC - [www.jisc.ac.uk](http://www.jisc.ac.uk) and Royal Mail - [www.royalmail.com](http://www.royalmail.com)).

It was not possible to make a comparison between the individual levels of conformance as there were so few

websites that conformed to Level AA or Level AAA (see section Websites and tasks).

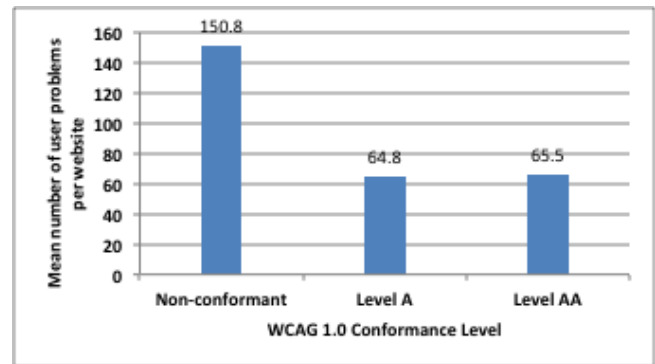


Figure 1. Mean number of instances of user problems per website grouped by WCAG 1.0 conformance levels.

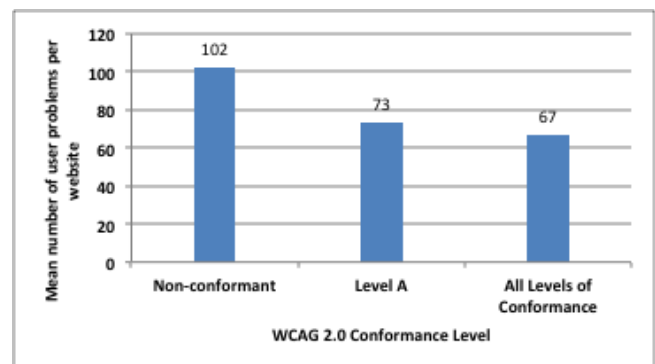


Figure 2. Mean number of instances of user problems per website grouped by WCAG 2.0 conformance levels.

A one-way ANOVA between non-conformant websites and Level A conformant websites showed no significant difference in the mean number of problems ( $F = 1.107$   $df = 1, 12$ , n.s.). One would expect a decrease in the number of user problems between non-conformant and Level A websites. A one-way ANOVA between non-conformant websites and websites that conformed at any level of WCAG 2.0 also failed to show a significant difference ( $F = 2.351$   $df = 1, 14$ , n.s.). Again, one would expect a decrease in the number of problems between non-conformant websites and conformant websites. In these analyses the three websites which failed SC 3.3.1 on one occasion only were classified as Level A (see section Websites and tasks).

### User problems and WCAG CPs/SCs violated

Analyses were performed on how the number of user problems relate to the number of CPs/SCs violated and the number of instances of CPs/SCs violated. For the following analyses, the TicketMaster website was omitted as the number of problems per user was more than 2 standard deviations above the mean for all the websites making it an outlier.

For WCAG 1.0, there was a significant correlation between the number of CPs violated and the mean number of problems per website per user ( $r = 0.53$ ,  $df = 14$ ,  $p < 0.05$ ).

For WCAG 2.0, there was also a significant correlation between the number of SCs violated and the mean number of problems per website per user ( $r = 0.54$ ,  $df = 14$ ,  $p < 0.05$ ).

For the instances of violations of CPs/SCs, there was no significant correlation between the instances of violations of WCAG 1.0 CPs or WCAG 2.0 SCs and the mean number of problems per website per user.

**User problems and coverage by WCAG CPs/SCs**

Analyses were performed on the extent to which user problems are covered by WCAG 1.0 CPs and WCAG 2.0 SCs.

For WCAG 1.0, Figure 3 shows the breakdown of user problems into categories of relevance of CPs and whether those CPs have been implemented on the website where the problem was encountered. The total percentage of user problems that were covered by CPs was 43.3% (the sum of bars 2 and 3 in Figure 3) and only a small percentage of those were implemented by developers (5.7% of all user problems, bar 3 in Figure 3, or 13.3% of all user problems covered by WCAG 1.0). This means that of the problems encountered by users on websites, well over half (57.1%) were not covered by WCAG 1.0.

For WCAG 2.0, Figure 4 shows a similar breakdown of user problems into categories of relevance and implementation of SCs. The total percentage of user problems that were covered by SCs was 50.4% (the sum of bars 2 and 3 in Figure 4) and a similarly small percentage of these were implemented by developers (8.4% of all user problems, bar 3 in Figure 4, or 16.7% of all user problems with relevant SCs). This means that for WCAG 2.0, the current set of guidelines for web accessibility, almost half of the problems encountered by users on websites are not covered.

A Related Samples Wilcoxon Signed Rank Test showed there was no significant difference in the coverage of user problems between WCAG 1.0 and WCAG 2.0 across the five relevance and implementation categories ( $W_+ = 1.5$ ,  $df = 1$ ,  $p = 1.0$ ).

**User Problems not covered by WCAG 2.0 SCs**

Analyses were performed to determine what categories of user problems WCAG 2.0 SCs do not cover well. As there was no difference in the coverage of user problems between WCAG 1.0 and WCAG 2.0, the analysis for WCAG 1.0 is not presented here but is broadly similar in results.

Table 2 presents user problem categories where WCAG 2.0 SCs did not cover at least 10 problems.

Of the six categories not covered by WCAG 2.0 SCs, there were two categories that accounted for 13.5% of all user problems. *Content not found in pages where expected by users* describes problems where users confidently followed a link to a page, but a piece of information that they expected to find there was missing. For example, on a

museum website, users followed a link to an object in the museum collection but did not find any information about the room in which that object is displayed, which they expected.

The category *Content found in pages where not expected by users* describes the inverse situation. For this category of problems, users eventually found the information they were looking for, but not where they expected it and not by a logical process of following the navigation provided by the website. For example, on the Automobile Association website (theAA.com), users were looking for driving tips. They did find this information under the link "Learn to Drive", but they were surprised that they were on such a page which did not match their mental model of the information architecture of such a site.

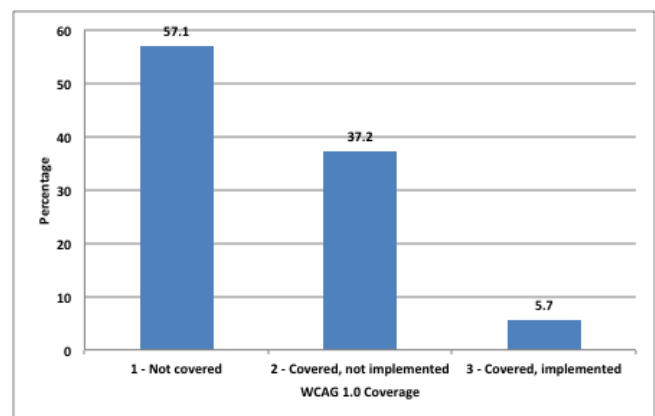


Figure 3: Categories of user problems divided by relevance of WCAG 1.0 CPs and implementation.

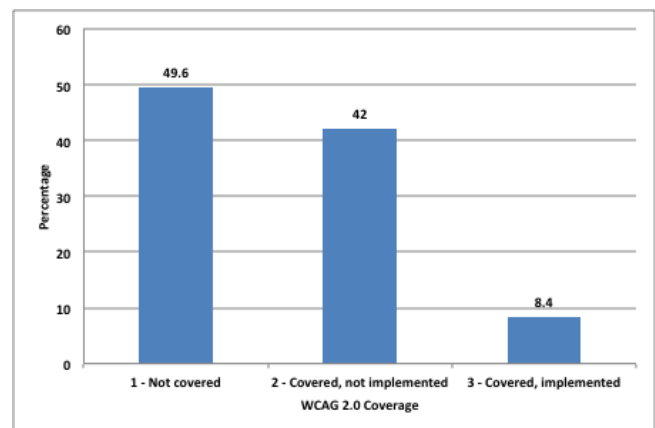


Figure 4: Categories of user problems divided by relevance of WCAG 2.0 SCs and implementation

Other categories had only a small percentage of their problems covered by WCAG 2.0. In the category *Irrelevant content before task content*, users often encountered problems with large blocks of content irrelevant to their task occurring before the relevant content.

Category	Total User Problems	% (No.) Covered by WCAG 2.0
Content found in pages where not expected by users	99	0
Content not found in pages where expected by users	88	0
Pages too slow to load	27	0
No alternative to document format (e.g. PDF)	17	0
Information architecture too complex (e.g. too many steps to find pages)	15	0
Broken links	10	0
Functionality does not work (as expected)	50	26.0 (13)
Expected functionality not present	29	34.5 (10)
Organisation of content is inconsistent with web conventions/common sense	39	35.9 (14)
Irrelevant content before task content	86	45.3 (39)
Users cannot make sense of content	65	67.6 (44)
No/insufficient feedback to inform that actions has had an effect	72	68.1 (49)

**Table 2. Categories of user problems with total number of problems and percentage (number) of user problems covered by WCAG 2.0 SCs.**

For example, when users were seeking information about insurance plans, the relevant page had lengthy descriptions of why it was important to buy insurance before a summary of insurance plans, the relevant content on the page. While there is an SC (2.4.1) that addresses skipping blocks of content repeated on multiple pages (e.g. main menu bars), there is nothing in this SC, or any other SC, describing the types of problems associated with irrelevant content that is unique to a single page.

**User problems covered by WCAG 2.0 SCs**

In order to understand why WCAG 2.0 SCs do not solve some problems encountered by users, analyses were performed on problems encountered on webpages where relevant SCs were implemented and yet users still had problems. Table 3 presents the categories of problems where more than 20% of the user problems in the category met these criteria of SC implementation and users having problems.

The largest number of user problems in Table 3 are in the category *Link destination not clear*, which accounted for 8.5% of all problems encountered by users. In 35.0% of

problems in this category, the website had properly implemented SC 2.4.4 regarding the description of link purpose, and yet users still had problems determining where the links lead. This indicates that the sufficient techniques for this SC, which are primarily aimed at addressing the problems of blind, screenreader users, are in fact not sufficient.

Category description	Total User problems	% (No.) Covered by WCAG 2.0 and implemented
Language too complicated for perceived target audience	9	33.3 (3)
Link destination not clear	117	35.0 (41)
Difficult to scan pages for specific items	8	37.5 (3)
No enhancements to multimedia content	31	51.6 (16)
Meaning in content is lost or modified due to transformations	6	83.3 (5)
No alternative to information presented in tables	12	100.0 (12)
Heading structure violated	9	100.0 (9)

**Table 3. Categories of user problems with their total number of problems and percentage (number) of user problems covered by WCAG 2.0 SCs and implemented.**

The category with the second largest number of user problems in Table 3 concerns enhancing multimedia with audio description. Audio description is an enhancement for multimedia where an additional audio track that describes what is happening in the video is played along with the original audio tracks to provide blind viewers with descriptions of vital visual information [12]. For a page containing pre-recorded videos, it passes SC 1.2.3 (Level A), if it provides an audio description or another alternative for all these videos. The only other alternative mentioned by WCAG is a text description of the videos with a text transcript of the audio tracks, all indexed by time. However, in a somewhat complex relationship between SC 1.2.3 and SC 1.2.5, if audio description (as opposed to the text description) is provided for all pre-recorded videos, the page also passes SC 1.2.5 (Level AA). As shown in Table 3, there were 31 problems in the category *No enhancements to multimedia content: audio description*. Of those problems, 51.6% of the web pages passed SC 1.2.3 at Level A by providing an appropriate text description. These problems were covered by WCAG 2.0 and were implemented correctly, but users rejected that implementation because they wanted an audio description. The remaining problems in the category were covered by

WCAG 2.0, but not implemented properly because there was no audio description or any other alternative provided.

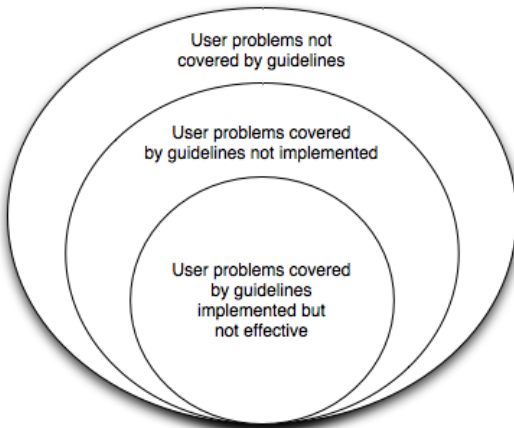
**DISCUSSION AND CONCLUSIONS**

This evaluation of 16 websites by 32 users has revealed a complex relationship between problems encountered by blind, screenreader users and WCAG 2.0.

Firstly, for WCAG 1.0, there was a significant decrease in the mean number of user problems between non-conformant websites and Level A conformant websites. The same was true when non-conformant websites were compared to Level AA websites.

However, for WCAG 2.0, there was no significant decrease in the mean number of user problems when comparing non-conformant websites and Level A conformant websites. There were so few websites that conformed to Level AA and Level AAA, similar tests could not be performed for those conformance levels. However, the same was true when non-conformant websites were compared to websites of all conformance levels.

These findings are quite unexpected. It seems that the upgrade to WCAG 2.0 has not had the expected effect. For WCAG 2.0, one would expect there to be a larger decrease in the number of user problems from non-conformant websites to Level A conformant websites than there was for WCAG 1.0. However, the results show conformance of a website to WCAG 2.0 Level A does not mean that users will encounter fewer problems on it and as a result it does not necessarily mean that following WCAG 2.0 will “make content accessible to a wider range of people with disabilities”[6].



**Figure 5: The overall set of user problems divided into three types: problems not covered by guidelines, those covered by guidelines but the guidelines are not implemented and those covered by guidelines with guideline implementations.**

However, for WCAG 2.0 there was a correlation between the number of problems encountered by users and the number of SCs violated. These results indicate that the current WCAG 2.0 priority levels are too crude of an accessibility measure.

The results showed three different types of problems encountered by users. These types of problems are presented in Figure 5.

When user problems were compared to WCAG 1.0, 57.1% did not have a CP that could be clearly identified as being *directly relevant*. For WCAG 2.0, only 49.6% of problems were addressed by *directly relevant* SCs. This means only half of the problems encountered by users were *covered*, thus the title of this paper. The move from WCAG 1.0 to WCAG 2.0 has not increased the coverage of user problems, as one would have expected. For those problems covered by WCAG 2.0, only 16.7% of the directly relevant SCs are being implemented on websites. This is a serious problem for three reasons. First, it indicates that web developers still struggle with creating accessible websites, possibly because their understanding of the guidelines is low or because of a lack of tool support. Second, for those SCs not implemented, it is not possible at this time to determine whether the user problems would be addressed by implementing the directly relevant SCs. Finally, for those SCs that were implemented, the implementations failed to solve the user problems. This shows that proposed implementations for solving accessibility problems *must* be evaluated with disabled users.

The results showed that blind users reported problems when they encountered unexpected content or when they could not find content on a website. WCAG 2.0 does not cover these problems. Some may assert that these are not accessibility problems, but instead are usability problems and do not need to be addressed in WCAG 2.0. The authors disagree with this assertion for the following reasons. First, web accessibility is about ensuring that people with disabilities can use the Web. In order for this to be achieved, we must address *all* of the problems that disabled users encounter on web pages.

Second, previous research has shown many problems are shared by blind users and mainstream users [25]. In that research, blind users reported significantly higher severity ratings than their mainstream peers for shared problems. That result makes it critical that these shared problems are solved. The information architecture problems described above would likely impact both disabled users and mainstream users.

Finally, WCAG 2.0 already contains a number of Guidelines and SCs that relate to usability problems, such as providing proper feedback and helping users identify errors. All of these points support the inclusion of a broader range of problems in WCAG 2.0.

The results of this study indicate that it is time to move away from the problem-based paradigm for web accessibility, where our primary goal is to eliminate problems encountered by users. Taking a lesson from usability research, web accessibility research must define a much broader set of design principles, based on user data,

that focuses on the use of the web by people with disabilities – not just on the problems they encounter. Once those design principles are clearly understood, only then can we look at proposing rules and heuristics that web developers can apply to evaluate their success in creating websites that people with disabilities can use well. This new paradigm will help us to discover the second half of the accessibility story.

#### ACKNOWLEDGMENTS

We would like to thank all the participants who took part in this study for their time and valuable input. Author André P. Freire was supported by CNPq – Brazilian Ministry of Science, grant 200859/2008-0.

#### REFERENCES

1. *Research Based Web Design & Usability Guidelines*. U.S. General Services Administration, 2004.
2. *Equality Act 2010*. Government Equalities Office, 2010.
3. Alonso, F., Fuertes, J. L., González, Á. L. and Martínez, L. On the testability of WCAG 2.0 for beginners. In *Proc. of W4A'10*, ACM (2010).
4. Babu, R. and Singh, R. (2009) Evaluation of Web Accessibility and Usability from Blind User's Perspective: The Context of Online Assessment, in *(AISeL)*, A. E. L., ed. *Americas Conference on Information Systems (AMCIS) 2009*, San Francisco, Paper 623, 2009.
5. Brajnik, G., Yesilada, Y. and Harper, S. Testability and validity of WCAG 2.0: the expertise effect. In *Proc. of ASSETS 2010*, ACM (2010), 43-50.
6. Caldwell, B., Cooper, M., Guarino Reid, L. and Vanderheiden, G. *Web Content Accessibility Guidelines (WCAG) 2.0*. Web Accessibility Initiative (WAI), World Wide Web Consortium (W3C), 2008. Retrieved from <http://www.w3.org/TR/WCAG20/> on Jan. 15, 2012.
7. Colwell, C. and Petrie, H. Evaluation of guidelines for designing accessible Web content. *SIGCAPH Comput. Phys. Handicap.*, 70 (2001), 11-13.
8. Cooper, M., Guarino Reid, L., Vanderheiden, G. and Caldwell, B. *Understanding WCAG 2.0: A guide to understanding and implementing Web Content Accessibility Guidelines 2.0*. Web Accessibility Initiative (WAI), World Wide Web Consortium (W3C), 2010. Retrieved from <http://www.w3.org/TR/UNDERSTANDING-WCAG20/> on Jan. 15, 2012.
9. Cooper, M., Guarino Reid, L., Vanderheiden, G. and Caldwell, B. *Techniques for WCAG 2.0: Techniques and Failures for Web Content Accessibility Guidelines 2.0*. Web Accessibility Initiative (WAI), World Wide Web Consortium (W3C), 2010. Retrieved from <http://www.w3.org/TR/WCAG20-TECHS/> on Jan. 15, 2012.
10. Disability Rights Commission. *The Web: access and inclusion for disabled people - A formal Investigation conducted by the Disability Rights Commission*. Disability Rights Commission, 2004.
11. Donnelly, A. and Magennis, M. Making Accessibility Guidelines Usable. In *Proc. Universal Access: Theoretical Perspectives, Practice and Experience*, Springer (2003), 56-57.
12. Freed, F. and Rothberg, M. *Accessible Digital Media Guidelines*, National Centre for Accessible Media Website, (2006). Retrieved from [http://ncam.wgbh.org/invent\\_build/web\\_multimedia/accessible-digital-media-guide](http://ncam.wgbh.org/invent_build/web_multimedia/accessible-digital-media-guide) on Jan. 15, 2012.
13. Goette, T., Collier, C. and Daniels White, J. An exploratory study of the accessibility of state government Web sites. *Universal Access in the Information Society*, 5, 1 (2006), 41-50.
14. Hackett, S. and Parmanto, B. A longitudinal evaluation of accessibility: higher education web sites. *Internet Research*, 15, 3 (2005), 281-294.
15. Henry, S. L. *Introduction to Web Accessibility*. Web Accessibility Initiative (WAI), World Wide Web Consortium (W3C), 2005. Retrieved from: <http://www.w3.org/WAI/intro/accessibility.php> on Jan. 15, 2012.
16. Kane, S. K., Shulman, J. A., Shockley, T. J. and Ladner, R. E. A web accessibility report card for top international university web sites. In *Proc. of W4A'07*, ACM (2007), 148-156.
17. Kelly, B., Sloan, D., Phipps, L., Petrie, H. and Hamilton, F. Forcing standardization or accommodating diversity?: a framework for applying the WCAG in the real world. In *Proc. of W4A'05*, ACM (2005), 46-54.
18. Lazar, J., Dudley-Sponaugle, A. and Greenidge, K.-D. Improving web accessibility: a study of webmaster perceptions. *Computers in Human Behavior*, 20, 2 (2004), 269-288.
19. Lazar, J. and Greenidge, K.-D. One year older, but not necessarily wiser: an evaluation of homepage accessibility problems over time. *Universal Access in the Information Society*, 4, 4 (2006), 285-291.
20. Leporini, B. and Paternò, F. Applying Web Usability Criteria for Vision-Impaired Users: Does It Really Improve Task Performance? *International Journal of Human-Computer Interaction*, 24, 1 (2008), 17-47.
21. Leuthold, S., Bargas-Avila, J. A. and Opwis, K. Beyond web content accessibility guidelines: Design of enhanced text user interfaces for blind internet users. *International Journal of Human-Computer Studies*, 66, 4 (2008), 257-270.
22. Lopes, R., Gomes, D. and Carriço, L. Web Not For All: A Large Scale Study of Web Accessibility. In *Proc. of W4A'10*, ACM (2010), 10.
23. Nielsen, J. *Usability Engineering*. Morgan Kaufmann, Boston, MA, 1993.
24. Petrie, H., Badani, A. and Bhalla, A. Sex, lies and web accessibility: the use of accessibility logos and

- statements on e-commerce and financial websites. In *Proc. of ADDW 2005*, University of Dundee (2005).
25. Petrie, H. and Kheir, O. The relationship between accessibility and usability of websites. In *Proc. of CHI'07*, ACM (2007), 397-406.
  26. Petrie, H., Power, C., Swallow, D., Velasco, C. A., Gallagher, B., Magennis, M., Murphy, E., Collin, S. and Down, K. The value chain of web accessibility: challenges and opportunities. In *Proc. of ADDW 2011*, Sun SITE Central Europe (2011).
  27. Power, C., Petrie, H., Freire, A. and Swallow, D. Remote Evaluation of WCAG 2.0 Techniques by Web Users with Visual Disabilities. In *Proc. of UAHCI: Design for All and eInclusion*, 6765, (2011), 285-294.
  28. Regan, B. Accessibility and design: a failure of the imagination. In *Proc. of W4A'04*, ACM (2004), 29-37.
  29. Rømen, D. and Svanæs, D. Evaluating web site accessibility: validating the WAI guidelines through usability testing with disabled users. In *Proc. of the 5th NordiCHI 2008*, ACM (2008), 535-538.
  30. Sloan, D., Gregor, P., Booth, P. and Gibson, L. Auditing accessibility of UK Higher Education web sites. *Interacting with Computers*, 14, 4 (2002), 313-325.
  31. Thatcher, J., Burkes, M. R., Heilmann, C., Henry, S. L., Kirkpatrick, A., Lauke, P. H., Lawson, B., Regan, B., Rutter, R., Urban, M. and Waddel, C. D. *Web accessibility: web standards and regulatory compliance*. Friends of ED, 2006.
  32. Vanderheiden, G., Chisholm, W. and Jacobs, I. *Web Content Accessibility Guidelines 1.0*. Web Accessibility Initiative (WAI), World Wide Web Consortium (W3C), 1999. Retrieved from: <http://www.w3.org/TR/WAI-WEBCONTENT/> on Jan. 15, 2012.
  33. Williams, R. and Rattray, R. An assessment of Web accessibility of UK accountancy firms. *Managerial Auditing Journal*, 18, 9 (2003), 710-716.