Progress on Website Accessibility?

VICKI L. HANSON, University of Dundee
JOHN T. RICHARDS, IBM T. J. Watson Research Center and University of Dundee

Over 100 top-traffic and government websites from the United States and United Kingdom were examined for evidence of changes on accessibility indicators over the 14-year period from 1999 to 2012, the longest period studied to date. Automated analyses of WCAG 2.0 Level A Success Criteria found high percentages of violations overall. Unlike more circumscribed studies, however, these sites exhibited improvements over the years on a number of accessibility indicators, with government sites being less likely than topsites to have accessibility violations. Examination of the causes of success and failure suggests that improving accessibility may be due, in part, to changes in website technologies and coding practices rather than a focus on accessibility per se.

Categories and Subject Descriptors: H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Input devices and strategies*; K.4.2 [**Computers and Society**]: Social Issues—*Assistive technologies for persons with disabilities*; K.4.2 [**Computers and Society**]: Social Issues—*Handicapped persons/special needs*; K.5.2 [**Legal Aspects of Computing**]: Governmental Issues—*Regulation*

General Terms: Standardization, Legal Aspects

Additional Key Words and Phrases: Web accessibility, disability

ACM Reference Format:
Hanson, V. L. and Richards, J. T. 2013. Progress on website accessibility? *ACM Trans. Web* 7, 1, Article 2 (March 2013), 30 pages.
DOI: http://dx.doi.org/10.1145/2435215.2435217

1. INTRODUCTION

As the economics of service and information provision drive more content exclusively to the Web, a large number of people with disabilities are disadvantaged. Once considered a luxury, access to the Web is now becoming a requirement for full participation in society. Consider three recent examples: Tickets to the 2012 Olympics were only available online; a recent UK art competition was only available to those who submitted their work electronically; top sources of news and information (e.g., The Huffington Post) and retail shopping (e.g., Amazon) are only available on the Web. These examples suggest that an accessible Web is becoming an imperative for a digital economy that includes all [Hanson et al. 2009]. Moreover, disabled individuals are not a niche sector of the population; worldwide, the estimated number of individuals with a disability is approximately 15% of the population [World Health Organization (WHO) 2012] and it has been estimated that more than half of the computer using population (which

This work was supported by a Royal Society Wolfson Merit Award WM080040 to V. L. Hanson and by RCUK EP/G066091/1 "RCUK Hub: Inclusion in the Digital Economy" and EPSRC grant EP/H006834/1.
Authors' addresses: V. L. Hanson (corresponding author), School of Computing, University of Dundee, Dundee, Scotland; email: vlh@acm.org; J. T. Richards, IBM T.J. Watson Research Center, P.O. Box 704, Yorktown Heights NY 10598.

DOI: http://dx.doi.org/10.1145/2435215.2435217

includes people with age-related disabilities) could benefit from accessibility features [Forrester 2003].

The importance of accessibility is highlighted in the guidelines for accessibility created by the W3C's Web Accessibility Initiative [W3C 2011a] and by the number of accessibility requirements created by governments [W3C 2006b]. Web authoring tools now include accessibility support, and computer science educators are creating accessibility courses and building accessibility into a more general curriculum [ACM Curricula Recommendations 2012; Ludi 2007; Poor et al. 2012; Waller et al. 2009].

Given these guidelines, regulations, tools, and educational initiatives, and the reported benefits of making websites accessible [W3C 2010a], it might be expected that access for disabled users would be improving. This is not clear, however. Reports over the years from many countries and across various types of sites (including commercial, government, news, healthcare, and higher education sites) on several accessibility criteria suggest low conformance with the criteria. The present study investigates indicators of accessibility for a relatively large number of websites over many years. Examining a corpus of U.S. and U.K. high traffic volume (topsites) and government sites, we explore success on machine-testable indicators over the 14 years from 1999 to 2012.

This work provides two main contributions to the literature. The first is the finding of some accessibility improvements over the 14 years studied. Specifically, there were fewer violations on supplying appropriate alternative text for nontext content, providing means of bypassing content, indicating the language for pages, and in using generally robust HTML/XHTML code. In most cases, there were fewer violations for government sites than topsites. Despite these positive trends, however, site compliance even on machine testable and relatively easily implemented aspects of Web accessibly is nowhere close to what it should be.

The second contribution of this work lies in uncovering some possible reasons for changes in accessibility. Analyses revealed, for example, that there were relatively few violations on some indicators that related to good coding practice. With our analyses of these websites, we begin to examine underlying reasons for both success and failure. By exploring the underlying drivers of Web accessibility we hope to gain insights that could help content, tool, standard, and guideline developers be more effective.

2. BACKGROUND

To put this research in perspective, we begin with a consideration of disability and with a discussion of Web accessibility guidelines.

The World Health Organization (WHO) defines disability in relation to the three factors of impairment, activity limitation, and societal participation. This is an important perspective, situating disability in relation to one's ability to engage in everyday tasks. Disability in terms of being able to use the Web can arise from a number of impairments. The Web Accessibility Initiative (WAI) of the W3C [2011a] has created guidelines designed to aid a broad population of disabled users, addressing, for example, the needs of people who are blind and need audio renderings of pages, those who have vision impairments requiring magnification or other changes to visual presentation, those who have hearing loss and need audio content visually displayed, those who have physical disabilities requiring alternative input (mouse, keyboard, or voice), and those who have cognitive or neurological disabilities such as dyslexia, seizures, or intellectual disabilities that require changes to page renderings or different means of generating input. A recent study by the WAI also looked at specific requirements of older Web users [W3C 2010b]. While other guidelines have been developed by government bodies and usability experts [Kelly et al. 2005], the WAI guidelines remain the authoritative worldwide reference.

The WAI has developed three sets of guidelines [W3C 2011a]. Garnering arguably the most research attention is the Web Content Accessibility Guidelines (WCAG). Intended to be used by website developers, these guidelines specify techniques for creating content that will be accessible for disabled individuals either directly or through the use of assistive devices. The best-known examples of assistive devices are screen readers. Intended primarily for users with severe visual impairments, screen readers require specific information on Web page structure in order to be able to make complex page layouts understandable using software that renders Web pages into speech. Other examples of assistive technologies include screen magnifiers, specialized mouse and joystick controls, and scanning single-switch access devices.

The second set of WAI guidelines are those that deal with user agents. These User Agent Accessibility Guidelines (UAAG) address how browsers, media players, and other devices and software that render Web pages should be created so that content will be accessible to disabled users. For example, the ability to change font sizes, now a capability of all major browsers, is a UAAG requirement that supports people with limited vision as well as many others with dyslexia or other cognitive disabilities who benefit from enlarged text or reduced information on a single screen.

The third set of guidelines is designed for those who create Web authoring tools. The goal of the Authoring Tool Accessibility Guidelines (ATAG) is for these tools to include the needed features for authors (able bodied or disabled) to create accessible content.

For the present research, the focus is on the WCAG guidelines. WCAG 1.0 was published in May 1999. For these guidelines, severity levels were assigned, with checkpoints being classified as priority 1, 2, or 3. Failure to comply with priority-1 checkpoints would be the most critical since this would mean that some groups of disabled users would be completely unable to use that portion of the website. Soon after the publication of these guidelines, reports began emerging that websites were not meeting standards, even for priority-1 checkpoints. These studies reported accessibility problems on websites from countries worldwide [Disability Rights Commission 2004; Ellison 2004; Hong et al. 2007; Lazar et al. 2003; Marincu and McMullin 2004] and across sites as varied as commercial, government, healthcare, culture, and higher education [Bigham et al. 2006; Davis 2002; Loiacono et al. 2009; Petrie et al. 2005; Zeng and Parmanto 2004].

Most of these studies examined a relatively brief time period. Two studies, however, examined accessibility over several years and thus are particularly relevant to understanding changes in Web accessibility over time. In the first, Hackett et al. [2004] conducted an analysis of the WCAG 1.0 guidelines, examining 22 U.S. government and 162 high-traffic nongovernment websites during the six years from 1997 to 2002. The same 22 government websites were examined in each of the 6 years. The non-government websites were not examined longitudinally, a different random sample being drawn instead each year from the top 500 site list. Based on 25 WCAG 1.0 checkpoints, Hackett et al. developed a Web accessibility score for each site's homepage based on the number of accessibility violations per page, weighted by criterion priority. Overall, they found that accessibility decreased over the course of the six years examined, due in part to the rapid development of new Web technologies not explicitly covered by the guidelines.

The second study that considered websites over multiple years examined accessibility of the homepages for the U.S. Fortune 100 corporate websites for the years 2000, 2003, 2004, and 2005 [Loiacono et al. 2009]. Using automated analyses for priority-1, -2, and -3 checkpoints, this research found no significant improvement in site accessibility for over the years studied (not all Fortune 100 sites were available for all years in the study, so this analysis does not directly compare the same sites over the years studied). Only 27% of the sites in the study were free of priority-1 violations; the number of

sites that met the priority-2 and priority-3 checkpoints was exceedingly low (ranging from 0 – 6 sites compliant on these two sets of checkpoints) during that timeframe.

Findings of poor accessibility compliance continued even following the introduction of government regulations. Lack of awareness about accessibility issues by those responsible for websites and a lack of clear guidance for developers are among the reasons cited for low rate of adoption [Kelly et al. 2005; Loiacono et al. 2009].

The WCAG 2.0 guidelines were published in December 2008. These updated guidelines were designed to address two specific challenges [W3C 2011b]. The first was rapid evolution and development of Web technologies. WCAG 2.0 was designed to be technology agnostic. Thus, rather than defining compliance with respect to particular technology, the guidelines considered user requirements organized into four areas, namely that Web pages must be perceivable, operable, understandable, and robust.

The second challenge was to create guidelines that were precisely testable for situations (such as procurement departments for government and business) where it was necessary to be able to specify and verify requirements. As part of the new guidelines, the underlying rationale for priority-1, -2, and -3 checkpoints was reconsidered as all criteria (even those in priority-3 checkpoints) can be critical for some disabled users. WCAG 2.0 therefore moved away from the notion of priority in terms of impact on users and instead focused on compliance levels. Level A is always required for a site to be accessible, while level AA and AAA are higher (more stringent) criteria. These levels suggest roughly an implementation order for content authors, with level A always being required. These orderings were based on a number of factors, such as possible workarounds and reasonable skills for authors to achieve [Reid and Snow-Weaver 2008].

The WCAG 2.0 guidelines provide 61 separate success criteria. Conformance requires that a page fully meet the success criteria for a specific level in order to be compliant at that level. Of the 61 success criteria, 25 are considered level A. These 25 level-A success criteria are shown in Table I.

In addition, WCAG 2.0 guidelines were developed such that all Success Criteria (SCs) would be testable, either automatically and with complete reliability (“machine testable”) or by knowledgeable humans (“human testable”) [W3C 2006a]. A number of automated accessibility checkers exist. Full compliance and usability testing, however, also requires human checking. Human testable criteria have a goal of 80% reliability, with 80% of knowledgeable human evaluators agreeing on how to meet criteria. Recent studies, however, have shown that this goal may be difficult to attain. Studies with both novice and experienced evaluators have found agreement far below the target [Alonso et al. 2010; Brajnik et al. 2012; Sampson-Wild 2007], with Brajnik et al. [2012] finding that evaluators reached this target only on about half of the success criteria.

Despite the efforts that have gone into website accessibility investigations, these various reports give a fragmented view of adherence to accessibility guidelines, with different sites, evaluation techniques, and criteria being examined in each study. While indicating accessibility problems, they do little to help us understand changes to accessibility implementations. They also do not add much to the previously discussed reasons for failures and do even less to help us understand why some criteria might be followed relatively successfully.

To tackle these issues, the present research examined whether accessibility indicators show improvement over the years since WCAG 1.0 was published, cataloging errors on a set of machine-testable, level-A success criteria. We examine WCAG criteria over 14 years, a considerably longer time period than any previous research. For these years, testing was conducted for both high-traffic websites (topsites) and government websites over two geographic regions. There is reason to believe that

Table I. WCAG 2.0 Principles, Guidelines, and Level-A Success Criteria

Principle	Guideline	Success Criteria
Perceivable	1.1 Text alternatives for non-text content	**1.1.1 Text alternatives for non-text content**
	1.2 Alternatives for time-based media	1.2.1 Pre-recorded audio and video only content 1.2.2 Captions (Subtitles) for pre-recorded audio content 1.2.3 Text or audio for pre-recorded video content
	1.3 Content presented in different ways	1.3.1 Information and relationships can be programmatically determined or available in text 1.3.2 If content sequence meaningful a correct reading sequence can be programmatically determined 1.3.3 Instructions do not rely on shape or sound alone
	1.4 Visual and audio elements accessible	1.4.1 Color alone is not used to convey information 1.4.2 Audio content is controllable by user
Operable	2.1 Keyboard accessible functionality	2.1.1 All page functionality should be available through the use of a keyboard 2.1.2 Keyboard focus is not locked or trapped
	2.2 Response time sufficient	2.2.1 Adjustable timings for applications 2.2.2 Provide means to pause, stop, or hide dynamically changing content
	2.3 Flashes below threshold	2.3.1 Flashing content conform to speed and number constraints
	2.4 User page navigation facilitated	**2.4.1 Provide means for bypassing blocks of content** **2.4.2 Provide page titles that describe topic or purpose** 2.4.3 Navigation order of focusable components makes sense 2.4.4 Purpose of links can be determined from link text or surrounding text
Understandable	3.1 Text readable and understandable	**3.1.1 Specify default human language of the page using the HTML language attribute**
	3.2 Pages operate in predictable ways	3.2.1 Change in component focus does not cause a change in context that would confuse a user 3.2.2 Entering data or selecting a form control has predictable effects
	3.3 User error correction	3.3.1 If an input error is detected, the item is identified and the error is described in text **3.3.2 Labels or instructions are provided for interactive elements**
Robust	4.1 Pages compatible with user agents and assistive technologies	**4.1.1 HTML/XHTML parsing errors are avoided** 4.1.2 Name, role, and value of all interface components can be programmatically determined

Machine-testable criteria, shown in bold, are examined in the present study.

topsites adopt new standards and recommendations more rapidly than small, less highly frequented sites [Chen 2009; Harper 2008]. Also, research with blind users has shown that such sites account for a disproportionately large amount of Web usage for

this population, with Google alone having been shown to account for 20% of the pages visited [Bigham et al. 2007].

The topsites and government sites we examined were from the United Kingdom (U.K.) and the United States (U.S.). Both have regulations that govern accessibility of government websites [W3C 2006b]. As such, both would be expected to be at the forefront of website accessibility. We analyzed a relatively large number of websites, 108 in total, over the 14 years of study, looking at conformance with key WCAG success criteria. At first blush it might seem strange to consider WCAG 2.0 criteria in the years before the publication of the guidelines. We note, however, that the success criteria tested were all embodied in WCAG 1.0. Our analyses of sites before WCAG 2.0 introduction allow us to see the extent to which these WCAG 1.0 accessibility features were applied before WCAG 2.0 and how the introduction of these new guidelines might have altered their adoption.

As we will see, aspects of our analyses question whether obtained accessibility improvements were due to guidelines per se, or whether other factors, not directly related to accessibility, might have indirectly created improvement in certain aspects of Web accessibility. In other work exploring how changes in coding practices over time might be influencing accessibility, we conducted a detailed manual inspection of four representative websites; two government, one popular news media, and one large-scale e-commerce website [Richards et al. 2012]. We found evidence that at least some accessibility improvements have resulted from an exploitation of new browser features to enhance page layout and design, an increased concern for page rank in search results, and changes in coding styles to attain both better cross-browser consistency and cross-device compatibility. In the current work, we again consider how changes in coding practices might be indirectly improving accessibility, examining the question within the context of our full corpus of topsites and government sites.

3. THE EXPERIMENTAL CORPUS

In selecting the corpus of websites for study, we were guided by the following.

(a) *Language.* All websites to be tested were to be sites from English-speaking countries. This facilitated the researchers' ability to understand the site and the developers' intentions.
(b) *Accessibility policy.* Websites to be tested should be from countries that have an existing accessibility policy related to the Web, based on WCAG standards.
(c) *Years.* The years to be studied include the years from which the WCAG 1.0 guidelines were released (1999) to the present.
(d) *Pages.* Only the top page from each website was analyzed. The reasoning for this was observation from our own and other work that top pages tend to represent the best accessibility for each site [Nielsen 2000]. As such, these pages represent arguably the most optimistic picture of accessibility.

Given the preceding criteria and the need to constrain the analysis to a reasonable number of websites, the top 60 sites in the UK and the US were selected for analysis. These sites were based on the Alexa ratings as of April 19, 2011 [Alexa 2011].

The URLs for the full list of these Alexa topsites are in the Appendix. The final count of UK topsites was reduced to 58 after removing lloydstsb.co.uk (duplicated and then superseded by www.lloydstsb.com) and pgmediaserve.com (for which no archived or current pages could be found).

Inspection of these lists of topsites revealed that there was only one government site in these lists, direct.gov.uk for the UK, which is a portal for a large number of government services. Arguably, the US Postal Service, another topsite, could be considered a government site, but it does not have .gov in its URL (usps.com) and is not actually

federally funded. Given the regulatory emphasis on (and importance of) government website accessibility, we therefore included as a second corpus an additional 10 sites for each country. These sites all had a .gov or .gov.uk in their URL. An attempt was made to find sites that were functionally equivalent for the two countries (such as the office of the Departments of Education for the two countries; both the UK Border Agency and the US Department of Homeland Security, and so on). Only the UK direct.gov.uk site appeared in the corpus both on the topsites list and the government list. The complete listing of the government sites selected appears at the bottom of the Appendix.

It is worth noting that there is considerable variability in the amount of content on these websites. While some are primarily search engines or portals to other sites (such as Google, Yahoo, and Bing), many are multifaceted sites with a rich variety of constantly changing information (such as Huffington Post, the Guardian, and BBC).

3.1. Archived Pages in Final Corpus

For each of the aforesaid Websites, the Wayback Machine Internet Archive [Wayback 2011] was used to retrieve the homepage of the site for the years 1999 through 2010. For each site, the first available instance of the homepage for each year (beginning with the January 1 calendar date) was used for our corpus. Homepages for 2011 and 2012 were captured directly from the sites on May 4, 2011, and January 1, 2012. Appendix A shows any missing years for a site. Sites could have been missing for a year as there was no archive entry for that URL for that year (most common) or the site name was clearly owned by a different entity and used for a different purpose than was true in 2011.

As can be seen from the list of topsites from the Appendix, there was some overlap in the topsites for the two countries. For example, wikipedia.org and youtube.com are the same for both. Therefore, for the corpus we considered the number of *unique topsites* to be the total of UK and US sites available for each year including only one instance of each duplicate site (such as wikipedia.org) for the two countries. The number of *government sites* is the combined total of the UK and US government sites available for the years sampled. The final corpus consisted of a total of 952 pages for the unique topsites and 231 pages for government sites over the 14 years examined. The number of unique topsites and government sites by year is given in Table II.

In considering Table II, one can view this corpus from two perspectives. One perspective is that of how many accessibility problems are encountered in any given year for users who might visit sites such as those in the corpus. The second perspective is how specific websites might have improved (or not) over the years studied. We consider our corpus from these two perspectives. When considering all 952 topsites and 231 government sites we refer to this as the *all sites* corpus. The final row in Table II shows the number of sites that had pages available for all years of the study (1999–2012). When considering the 34 topsites and 10 government sites that had pages in all years studied, we refer to this as the *trend sites* corpus. These trends sites were used when considering whether there were accessibility improvements over the years.

3.2. Accessibility Statements

A simple but interesting indicator of awareness of, and commitment to, accessibility can be based on the presence of accessibility statements on these top-level pages. Accessibility statements, while not required by WCAG, are meant to indicate claimed conformance level and can provide accessibility guidance for people who come to the site. We conducted an automated search over our corpus, checking whether each page had a link with the word “accessibility” on the page. Note that this and all remaining string comparisons in this article are case insensitive. Note, also, that the Wayback

Table II. Final Page Corpus for Both the Full Corpus of All Sites and the Corpus of Trend Sites (for which pages were available in all years from 1999–2012)

Year	# Unique Topsites	# Government Sites
1999	44	11
2000	52	12
2001	54	13
2002	55	15
2003	59	14
2004	66	16
2005	68	17
2006	72	17
2007	76	19
2008	76	19
2009	78	18
2010	76	20
2011	88	20
2012	88	20
Complete 1999–2012 (trends corpus)	34	10

Table III.

	Linked Accessibility Statements		Mention of Accessibility	
	Full Corpus	**Trends Corpus**	**Full Corpus**	**Trends Corpus**
Unique Top sites	12.48% (2.06)	12.78% (1.88)	14.03% (2.32)	14.71% (2.15)
Government Sites	47.69% (7.69)	35.71% (6.85)	50.92% (7.78)	42.14% (7.05)

Internet Archive's own content, automatically inserted around archived page content, was removed prior to this and all remaining analyses.

Shown in Table III is the percentage of all sites in the corpus that had links to accessibility statements. As is apparent, few of these topsites had an accessibility statement linked to their homepage. In contrast, such accessibility statement links were significantly more common for government sites, $t(14.91) = 4.49$, $p<.001$, two-tailed for unequal sample variances, $d = 2.33$[1].

A slightly more tolerant criterion is whether there is any mention of "accessibility" anywhere in the page. Conformance statements need not be embedded in a linked page but can take the form of textual statements or compliance logos (in which case we make the reasonable assumption that the alt attribute for a logo signifying accessibility conformance would include the word "accessibility"). Even with this relaxed criterion, there were more government sites than topsites that mentioned accessibility, $t(15.29) = 4.55$, $p < .001$, two-tailed for unequal sample variances, $d = 2.33$.

Looking at the trends over the 14 years studied, both accessibility statement links and mentions continued to be more common on government than topsites. Table III also shows the percentages for the trends corpus in terms of linked accessibility statements and mentions of accessibility. The government sites were more likely to include linked statements and mention accessibility than were the topsites [$t(14.96) = 3.23$, $p = .006$, $d = 1.67$ and $t(15.40) = 3.72$, $p = .002$, $d = 1.89$, respectively, both two-tailed for unequal sample variances]. The relatively low number of non-government pages

[1]Data were normally distributed for all t-tests. Unless otherwise indicated, the data for independent t-tests also passed the assumption for homogeneity of variance, $p>.05$, on Levene's test for equality of variances.

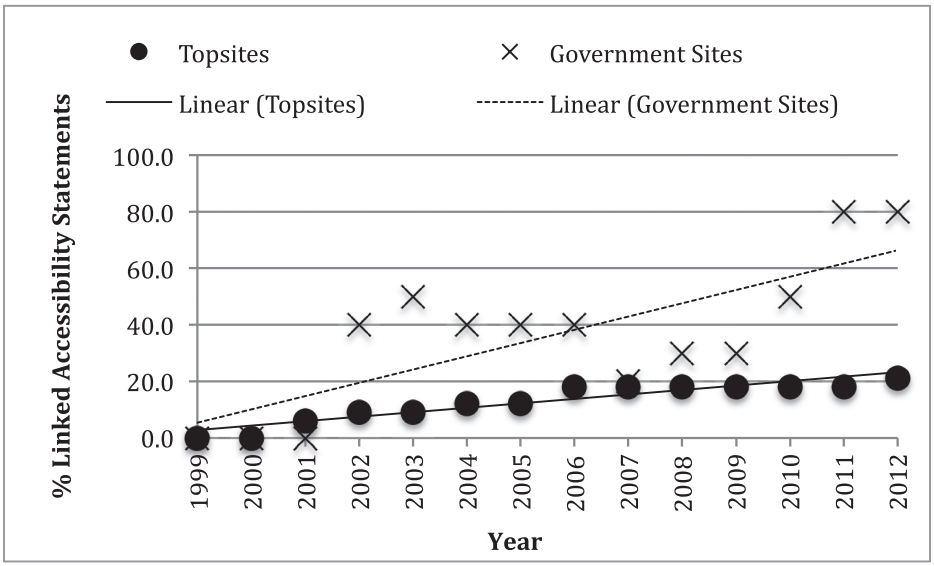

Fig. 1. Percentage linked accessibility statements and associated linear trends.

with accessibility statements should make us wonder about the degree of focus on accessibility.

As shown in Figure 1, the 34 topsites in the trends corpus exhibited only a small but significant increase in the use of accessibility links. In contrast, government sites exhibited a more striking increase in the use of these links. There was a significant correlation between year and the percentage of linked statements for both topsites, rho = .957, n = 14, $p < .001$, and government sites, rho = .665, n = 14, p = .009. As shown in Figure 2, the same pattern was found for accessibility mentions, again with significant correlations both for topsites, rho = .966, n = 14, $p < .001$, and government sites, rho = .755, n = 14, p = .002.

4. SUCCESS CRITERIA EVALUATION

Analyses will be grouped within each of the four WCAG 2.0 principles: Perceivability (WCAG 2.0 SC 1.x), Operability (WCAG 2.0 SC 2.x), Understandability (WCAG 2.0 SC 3.x), and Robustness (WCAG 2.0 SC4.x). The relationship to WCAG 1.0 will be noted in each case.

We limit our analyses to level-A success criteria. Level A is the minimal level required for a Website to be in conformance. Our evaluation of the corpus focused on machine-testable criteria, both because of tractability and relative lack of ambiguity. We note that not all level-A success criteria could be evaluated via (reasonably computable) automatic procedures. Some specific examples of criteria that could not be automatically evaluated include the following.

— Alternatives should be provided for time-based media (WCAG 2.0 SC 1.2): This guideline addresses the needs of those with hearing loss to have visual (text or sign language) alternatives presented for files such as video and podcasts that have audio content; it also addresses the need for those with visual impairments to have audio alternatives for video files.
— Color and other sensory characteristics should not be used as the sole means of conveying content (WCAG 2.0 SC 1.3.3 and 1.4.1): This addresses users who are

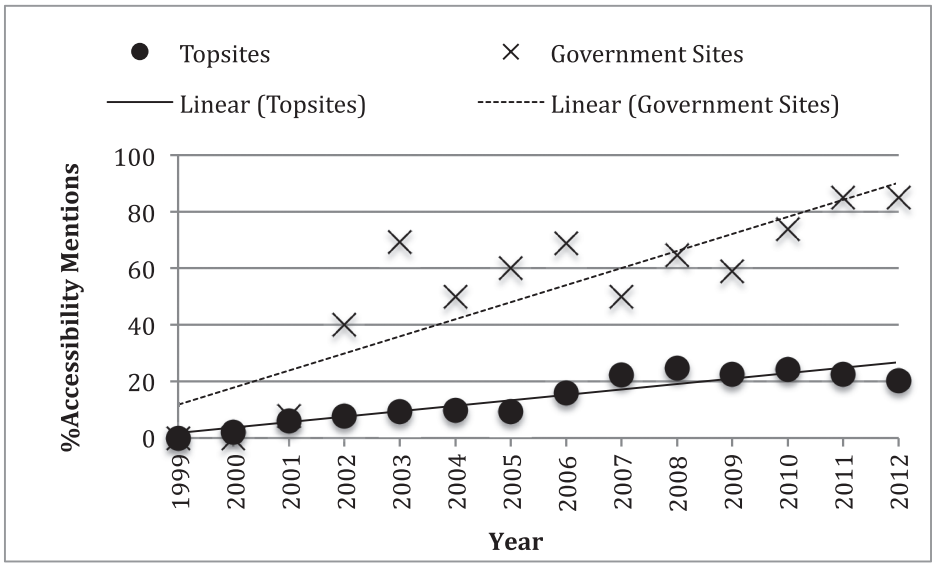

Fig. 2. Percentage accessibility mentions and associated linear trends.

visually impaired and focuses, for example, on the fact that statements such as "click the green button" or "wait for the beep" will not be accessible to some users.

— Flashing should be limited (WCAG 2.0 SC 2.3): This addresses users who are susceptible to seizures, requiring that video and any other visual content that has flashing sequences be limited so as not to induce such seizures.

The set of evaluation criteria examined in the present study are shown in bold in Table I.

4.1. Perceivability

The perceivability principle addresses the fact that Web users with vision or hearing loss will be unable to access some forms of content unless it can be changed into other perceivable forms.

4.1.1. Text Alternatives: Images. Arguably the most researched aspect of the WCAG guidelines is the inclusion of alternative descriptions of images, often referred to as the "alt tag", such as the following.

```
<img src="mysite.gif" height="60" width="100" alt="logo for
mysite.com" />
```

Complex images may also include a long description of their content using the longdesc attribute, such as what follows.

```
<img src="mysite.gif" height="60" width="100" alt="logo for
mysite.com" longdesc="mysite.html" />
```

Alternative descriptions of images were a priority-1 checkpoint in WCAG 1.0 and a level-A success criterion in WCAG 2.0, requiring that text alternatives be provided for all nontext content (SC 1.1.1). Without these alternative text descriptions, images will be inaccessible for users of screen readers as they cannot derive the meaning of images and must rely on the alternative text to provide information about images and other nontext media to users. This guideline further specifies that when nontext elements

Table IV. Percentage of Alt Tag Violations for Both Content Images (>32 pixels in both width and height) and Decorative Images (<8 pixels in either width or height)

	Image Type	
	Content Images	**Decorative Images**
Unique Topsites	42.0%, SE = 4.29 (8,429)	66.38%, SE = 5.02 (16,924)
Government sites	24.0%, SE = 4.29 (1,264)	60.9%, SE = 5.02 (2,834)

Shown are the mean percentages, SE, and the number of analyzed image tags in parentheses.

are for purely decorative or structural purposes (such as transparent or background-colored "spacers" used to position content) an `alt` attribute should be provided but with *null* content, that is, `alt=""`. Image tags with no `alt` attribute are read aloud by screen readers with a description such as "image" (which is meaningless for a content image and needlessly distracting for a purely decorative image).

Our corpus was analyzed for the number of violations in providing alternative text for images. We note that automated checkers do not detect if alternative text is semantically correct [Bigham et al. 2006; Borodin et al. 2010]. Thus, for example, if the text for an alt tag was "logo of the University of Dundee," there would be no check on whether this actually was the university logo.

For our automated analysis we did not attempt to consider semantic appropriateness. We did, however, set criteria to distinguish between content and decorative images. For our analysis, `img` tags were extracted from the page content and were sorted into probable content and decorative categories. Content images were defined as those having both `width` and `height` attributes greater than 32 pixels. Decorative images were those having either `width` or `height` less than 8 pixels. These two thresholds were chosen to maximize the number of images that were unambiguously either content or decorative. Ambiguous images with a width or height between these two thresholds were not included in further analyses. If neither width nor height were specified the image was classified as unknown and was also not included.

Correct alternative text for content images was defined as having an alt attribute with more than one character that was also not the string "image" or a string containing ".jpg", ".jpeg", ".gif", or ".png". Correct alternative text for decorative images was defined as having an `alt` attribute with null content (namely, `alt=""`). These criteria are more stringent than those generally applied by automatic accessibility tests but are more in keeping with the intent of the WCAG guidelines [Bigham et al. 2006].

Reported in Table IV are the percentages of alt text violations for content and decorative images for all sites in the corpus. An analysis of variance was carried out on the factor of site type (topsites, government sites) by the repeated measure of image type (content, decorative). Content images (M = 32.98%, SE = 3.44), overall, were less likely to have violations than were decorative images (M = 63.62% SE = 3.53), $F(1,26) = 102.35$, $p < .001$, $\eta_p^2 = .797$.

The alt tags on government sites (M = 42.43, SE = 4.15) did not have significantly fewer violations overall than did the topsites (M = 54.17, SE = 4.18), $F(1,26) = 4.00$, $p > .05$, $\eta_p^2 = .13$, although these two factors interacted, $F(1,26) = 4.23$, $p = .05$, $\eta_p^2 = .14$. A post hoc analysis showed that there was a significant difference between topsites and government sites for content images, $t(26) = 2.96$, $p < .01$, $d = 1.161$, but not for decorative images, $t(26) = .78$, $p > .05$.

Examining trends for the sites that had Web pages available for the 14 years studied, the number of alt tag violations tended to decline over the years for both topsites and government sites. Table V shows the mean percentage of alt tag violations by year. As indicated in the bottom row of Table V, there were negative correlations between year

Table V. Mean Percentage of Violations on Image Tag Alternative Text by Year

	Topsites		Government Sites	
Year	**Content Images**	**Decorative Images**	**Content Images**	**Decorative Images**
1999	63.2% (185)	98.9% (451)	46.2% (52)	95.5% (22)
2000	63.9% (216)	95.4% (348)	46.4% (84)	82.5% (63)
2001	74.5% (238)	91.1% (756)	14.7% (75)	81.7% (60)
2002	73.9% (245)	80.2% (1121)	24.1% (79)	42.9% (247)
2003	51.9% (291)	72.5% (1639)	8.4% (95)	68.6% (299)
2004	51.1% (284)	62.7% (1575)	16.7% (66)	66.7% (294)
2005	51.1% (266)	51.1% (266)	19.2% (52)	75.2% (326)
2006	46.6% (313)	49.5% (1226)	16.1% (62)	76.7% (374)
2007	44.7% (257)	39.3% (963)	10.6% (66)	39.2% (204)
2008	24.5% (335)	35.8% (271)	16.3% (49)	67.7% (31)
2009	17.6% (296)	23.4% (380)	16.7% (54)	73.3% (30)
2010	13.7% (358)	34.0% (235)	4.7% (64)	58.2% (12)
2011	22.6% (451)	45.8% (107)	14.9% (47)	50.0% (4)
2012	26.9% (439)	52.4% (84)	14.9% (67)	50.0% (4)
Mean	44.7%	59.5%	19.3%	66.3%
SE	5.46	3.30	6.54	4.37
rho	−.898**	−.504	−.829**	−.603*

Reported in the bottom row are correlations with year, all n = 14, * $p < .05$, ** $p < .005$, two-tailed.

and percentage violations reflecting this decline. These correlations were statistically significant except for decorative images on topsites.

Consistent with the results for all sites in the corpus, the results for these trend sites suggest fewer violations for content image tags than decorative image tags, $F(1,26) = 156.09$, $p < .001$, $\eta_p^2 = .857$, and an interaction of image type by site type, $F(1, 26) = 42.44$, $p < .001$, $\eta_p^2 = .620$. As for the full corpus, there was a significant difference between topsites and government sites for content images, $t(26) = 3.98$, $p < .01$, $d = 1.561$, but not for decorative images, $t(26) = .87$, $p > .05$.

Table V shows that there were significant negative correlations for content images, such that the number of violations decreased over the years studied. For decorative images, violations decreased for the government sites, but not the topsites.

As for all sites, there was not a significant difference in the percentage of violations for topsites and government sites, $F(1,26) = 1.90$, $p > .05$, $\eta_p^2 = .068$.

To gain insight into specific accessibility failures, we examined the nature of the violations. For content images there are five easily detectable ways in which the `alt` attribute could fail to meet the WCAG recommendations: no `alt` attribute, empty `alt` string, alt string is name of image file, alt string is only one character long, alt string is the word “image”.

We examined the data for each of these five types of errors. Of the 3,338 failures across our full corpus, most were of the first and second type, such that there was no alt attribute or an empty alt string. There were only 59, 68, and 3 instances of the latter three, respectively.

Examining the sites in the trend corpus for these first two error types, an interesting pattern was apparent. As shown in Table VI, there was a tendency for the percentage of missing alt tags to decrease over the years and the percentage of empty alt tag strings to increase. This was particularly apparent for government sites.

For decorative images, only two types of errors are possible: there could be no alt tag provided or the alt tag could be nonempty. As shown in Table VII, the relative

Table VI. Error Analysis for Content Images for the Trend (sites for which data was available for all 14 years)

	Unique Topsites			Government Sites		
Year	**Total**	**Missing alt tag**	**Empty alt tag**	**Total**	**Missing alt tag**	**Empty alt tag**
1999	116	60.5%	2.2%	24	46.2%	0%
2000	138	61.6%	2.3%	39	46.4%	0%
2001	174	69.3%	3.8%	11	14.7%	0%
2002	181	69.0%	4.9%	19	20.3%	3.8%
2003	150	46.0%	5.2%	8	4.2%	3.2%
2004	144	43.3%	7.0%	11	6.1%	9.1%
2005	137	42.5%	7.9%	10	5.8%	7.7%
2006	141	33.2%	11.2%	10	6.5%	6.5%
2007	112	16.0%	26.8%	7	3.0%	4.5%
2008	81	13.4%	10.1%	8	4.1%	8.2%
2009	54	4.7%	12.5%	9	0.0%	11.1%
2010	49	4.7%	8.9%	3	1.6%	3.1%
2011	98	7.3%	14.4%	7	2.1%	12.8%
2012	115	6.2%	20.0%	10	13.4%	1.5%
Mean		34.1%	9.8%		12.5%	5.1%
SE		6.71	1.86		4.12	1.12
rho		−.924**	.895**		−.732**	.528

Shown are the number of content images that had errors and the percentage of each of the two predominant error types. Reported in the bottom row are correlations with year, all n = 14, ** $p < .005$, two-tailed. Note that the two types of errors do not sum to 100% for topsites or government sites, as these are not the only types of errors.

Table VII. Error Analysis for Decorative Images

	Unique Topsites			Government Sites		
Year	**Total**	**Missing alt tag**	**Nonempty alt tag**	**Total**	**Missing alt tag**	**Nonempty alt tag**
1999	446	97.6%	1.3%	21	45.5%	50.0%
2000	332	93.4%	2.0%	52	14.3%	68.3%
2001	689	84.8%	6.3%	49	35.0%	46.7%
2002	899	74.1%	6.1%	106	21.9%	21.1%
2003	1188	61.6%	10.9%	205	12.4%	56.2%
2004	987	53.9%	8.8%	196	8.5%	58.2%
2005	631	43.4%	7.9%	245	16.9%	58.3%
2006	607	46.7%	2.8%	287	20.9%	55.9%
2007	378	37.3%	2.0%	80	1.0%	38.2%
2008	97	25.1%	10.7%	21	3.2%	64.5%
2009	89	13.9%	9.5%	22	0.0%	73.3%
2010	80	14.0%	20.0%	7	0.0%	58.3%
2011	49	25.2%	20.6%	2	0.0%	50.0%
2012	44	26.2%	26.2%	2	0.0%	50.0%
Mean		49.8%	8.2%		12.8%	53.5%
SE		7.67	1.67		3.81	3.46
rho		-921**	.664**		−.882**	.146

Shown are the number of decorative images that had errors and the percentage of each error type. Reported in the bottom row are correlations with year, all n = 14, ** $p < .005$, two-tailed. Note that the two types of errors do not sum to 100% for topsites or government sites, as these are not the only types of errors.

proportion of these error types changed over the years. For both topsites and government sites, the predominant error in 1999 was missing alt tags. By 2012, the predominant errors were those in which there was an alt tag, but the alt tag value was

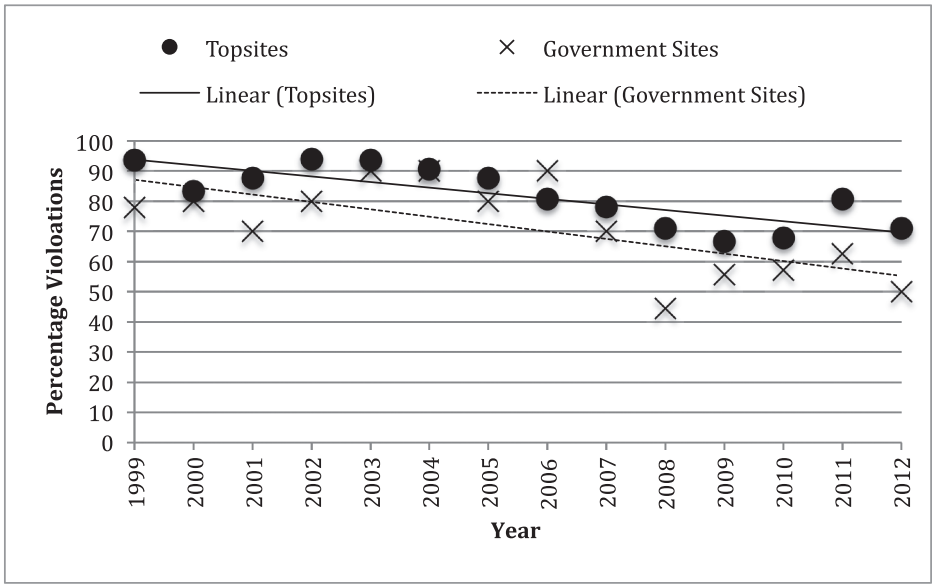

Fig. 3. The percentage of Websites that have violations in their use of alternative text for images over the years studied.

nonempty. This reversal is more striking in the case of government sites where all of the errors in 2012 were of the latter type.

The reasons for these patterns for content and decorative images are not clear, although the decrease in missing alt tags over the years suggests some attention to accessibility (either directly by content authors or indirectly by automatic tool insertion). The relative increase in empty alt strings for content images and nonempty strings for decorative images suggests a failure to properly understand or use the alt attribute.

While these results suggest that there is improvement in providing alternative text over the years, it needs to be kept in mind that conformance with WCAG 2.0 requires that *all* elements within a page be fully compliant. If we look at how well sites do on just this one aspect of the guidelines, we can see that there are still a number of violations. Figure 3 shows the percentage of government and topsites in the trend sites that had one or more violations in their use of homepage alt tags across both content and decorative images.

Although there were significant correlations over the years, such that the number of violations tended to decrease for both topsites, rho = $-.790$, n = 14, p = .001, two-tailed, and government sites, rho = $-.624$, n = 14, $p < .05$, two-tailed, it is apparent that few sites were free of accessibility violations on even this one criterion.

We now consider two additional cases in which guidelines specify the use of alternative text. For complex images, WCAG guidelines suggest the longdesc attribute be used. This attribute was not used in any of the government sites and was seldom used in any of the topsites in our full corpus, however. Perhaps more importantly, there is evidence that the use of this attribute (at least from an accessibility standpoint) is poorly understood, with it being used, for example, to give the URL of the gif of a blog contributor's headshot. In the full corpus of all sites, there was one instance of this attribute on the topsites in 2007, 45 in 2009 (all from the same website), 54 in 2010 (from two sites), 94 in 2011 (from one site), and 69 in 2012 (from one site). To summarize,

Table VIII. Form Input Violations for Image Buttons

Year	Unique Topsites	Government Sites
1999	75.0% (7)	-- (0)
2000	50.0% (10)	0.0% (3)
2001	50.0% (17)	0.0% (3)
2002	46.7% (27)	0.0% (3)
2003	47.1% (32)	0.0% (7)
2004	33.3% (40)	0.0% (8)
2005	42.9% (32)	0.0% (10)
2006	28.6% (32)	0.0% (10)
2007	38.9% (37)	0.0% (12)
2008	47.1% (42)	0.0% (5)
2009	42.9% (51)	0.0% (6)
2010	28.6% (46)	20.0% (5)
2011	22.2% (33)	20.0% (7)
2012	20.0% (22)	16.7% (10)
Mean	40.9%	4.36%
SE	3.78	2.31
rho	−.882**	--

Shown in parentheses is the number of image buttons for each year in the corpus. The percentage is the percentage of full site compliance for these image buttons. Reported in the bottom row are correlations with year, all n = 14, ** p<.005, two-tailed.

only three of the topsites used longdesc, with all except nine of the occurrences being from one site.

Applets were also very rare in our corpus. Over all sites in the corpus, the government sites had only seven instances of applets, all prior to 2005. Only two of these applets had alternative text. The topsites had only two applets, both prior to 2006. In neither case did the applet have the needed alternative text.

4.1.2. Text Alternatives: Image Buttons and Form Inputs. Also falling under WCAG's success criterion 1.1.1 are form input elements that require alternative text in order to be accessible. Image buttons, radio buttons, check boxes, text fields, etc., all require alternative text descriptors.

Consider, for example, image buttons. These buttons, a type of form input, often contain critical information for page usage (such as "Buy Now", "Create Account", "Register", or "Search"), as indicated in the following example.

```
<input type="image" width="41" height="31" border="0" src="
http://pics.ebay.com/aw/pics/homepage/v2/finditTop_41x41.gif"
alt="Find It!">
```

Without the required alternative text, the function associated with such a button will be difficult to discern for those using a screen reader. The number of image buttons was relatively small in our full corpus, especially for government sites. There were 952 image buttons on the topsites and only 231 on the government sites. With only a couple of exceptions, all image buttons for these government sites had a nonnull alt attribute (note that the stricter definition of accuracy for nonnull strings reported before, such as being longer than one character in length, was not used in this analysis). Only 3.2% of the government sites had an image button without an alternative text. For the full set of topsites, however, the use of a nonnull alt attribute was well below 100%. Moreover 46.4% of the topsites had at least one image button without alternative text.

Table IX. Percentage of Violations for Text, Checkbox, Radio, File, and Password Input Types, along with Textarea and Select Elements (combined), having an Explicit Label as Recommended by WCAG 2.0

Year	Unique Topsites	Government Sites
1999	100.0% (67)	100.0% (4)
2000	100.0% (74)	100.0% (7)
2001	100.0% (80)	100.0% (10)
2002	100.0% (91)	100.0% (9)
2003	89.7% (78)	45.5% (11)
2004	76.4% (110)	23.1% (13)
2005	74.8% (111)	37.5% (16)
2006	69.7% (132)	35.3% (17)
2007	64.0% (125)	31.6% (19)
2008	58.3% (103)	27.8% (18)
2009	61.4% (140)	33.3% (18)
2010	55.7% (131)	33.3% (18)
2011	42.1% (159)	29.4% (17)
2012	60.2% (113)	57.9% (19)
Mean	75.2%	53.9%
SE	5.24	8.38
rho	−.953**	−.614*

Shown in parentheses is the number of such controls in each year. Reported in the bottom row are correlations with year, all n = 14, *p < .05, ** p < .005, two-tailed.

Analysis of the sites in the trends corpus showed a tendency for topsites to have fewer errors over the years. As shown in Table VIII, there was a significant negative correlation between year and the number of errors for the topsites. Interestingly, while there were very few violations on the government sites, those that did occur were all present in the most recent three years.

In addition to alt tags on image buttons, label elements should be used to associate text labels with form controls. Looking at input elements of type text, checkbox, radio, file, and password, along with textarea and select elements, we find a similar tendency to improve over time.

Implicit labels (in which the input control is nested within the label tag) have not been reliably handled by assistive technologies and are not recommended (and are not counted as successful in our analysis). The percentage of these controls having labels in violation of WAI guidelines was not significantly higher for topsites than government sites, $p>.05$, when considering the full corpus.

Table IX shows the percentage of violations for these controls for our trends corpus. The number of controls in our corpus that should have labels is shown in parentheses. There were more violations for topsites than government sites, $t(21.83) = 2.15$, $p = .043$, two-tailed, $d = .92$. For both, there were significant negative correlations for the percentage of violations, indicating that violations tended to decrease over the years for both topsites and government sites.

In summarizing the data for alternative text, the situation remains fairly poor for disabled users who need such text to be able to make sense of images (and the occasional applet). There is evidence of improvement over the years, but much still needs to be done to attain full page compliance. Keeping in mind that alternative text (alt tags) is probably the best known accessibility requirement, is directly supported in authoring tools, and is fairly easily tested through automatic means, we are again left to

wonder if accessibility guidelines and government regulations are having the impact we might hope for.

4.2. Operable

The WCAG 2.0 principle for operability addresses the need for Web pages to provide sufficient time for input to be completed by users of assistive technologies, to avoiding flashing that could cause seizures, and to be navigable by people using diverse means (e.g., a keyboard instead of a mouse). In our automatic checking, we were able to reliably test two of the success criteria related to navigation.

4.2.1. Navigable: Bypass Blocks. The intent of this navigation guideline is to allow users to skip over (bypass) blocks of (relatively secondary) content. Sighted users can easily ignore such blocks based on visual inspection of the page layout. For users who rely on audio renderings of pages, however, this is not possible and it can take quite a bit of time to scan through text that is spoken prior to finding the content of interest. For sighted users who rely on keyboard-only access, these bypass blocks also allow for fewer keystrokes to reach material of interest. The ability to bypass blocks of material was a WCAG 1.0 priority-1 checkpoint and is a WCAG 2.0 level-A success criterion (SC 2.4.1).

There are several ways that this bypass can be accomplished by either creating links to explicitly skip to primary blocks or grouping blocks of material in a way that can be skipped. In the present work, we examined three of the ways: using "skip navigation," using headings, and using frames to group blocks of materials.

A correctly formed skip navigation element (when coded in the most typical fashion) has a link near the top of a page that usually goes directly to the main content. For this to work, there needs to be both an href link and a matching anchor, as shown next:

```
<a href="#main">Skip Navigation</a>
...
<a name="main"></a>
```

There was a very large difference between topsites and government sites in their use of skip navigation. Considering all sites in the corpus, use of skip navigation was much more common for government sites than topsites. Starting in 2002 the government sites began implementing skip navigation; since 2005 about 70% of the sites each year have had skip navigation links. Fewer than 20% of the topsites in the corpus adopted this over the years studied. This difference in the percentage of government (mean = 48.64%, SE = 27.89) and topsites (mean = 11.08%, SE = 7.41), was significant, $t(14.83) = 4.87$, $p<.001$, two-tailed, for unequal variances of the two samples, $d = 2.529$.

Looking at the trend sites corpus in Table X, we can see that starting in 2002 both government sites and topsites began implementing skip navigation which has become quite prevalent by 2012. Topsites, however, have been much less likely to adopt this. The difference between government and topsites was highly significant, $t(14.83) = 5.05$, $p<.001$, two-tailed, for unequal variances of the two samples, $d = 2.62$.

Closer examination of specific errors revealed evidence of a failure to understand proper technique (or at least incomplete testing). There were 951 pages between topsites and government sites that failed to have skip navigation links, most missing both the initial href and the target. Furthermore 19 had the initial href but were missing the target, two were missing the text indicating to the user that there was a means to skip to the content, one incorrectly coded the target name with a leading #, and one had a skip to the top of the page. This latter error was syntactically correct, but

Table X. Percentage of Sites in the Trends Corpus Using Skip Navigation

Year	Unique Topsites	Government Sites
1999	0.0%	0.0%
2000	0.0%	0.0%
2001	0.0%	0.0%
2002	3.0%	30.0%
2003	6.1%	40.0%
2004	15.2%	50.0%
2005	18.2%	60.0%
2006	21.2%	50.0%
2007	9.1%	60.0%
2008	18.2%	50.0%
2009	18.2%	60.0%
2010	21.2%	50.0%
2011	33.3%	60.0%
2012	30.3%	80.0%
Mean	13.86%	42.12%
SE	2.96	6.80
rho	.921**	.865**

Reported in the bottom row are correlations with year, all n = 14, ** $p < .005$, two-tailed.

apparently reflected a misunderstanding of the accessibility requirement of bypassing blocks of content.

Another means by which developers can meet the standard for bypassing blocks of material is to provide headings to structure page content, one that is the most commonly employed by screen reader users [Borodin et al. 2010]. It is quite difficult for automated testing to determine if a particular configuration of headings is sensible and affords efficient page navigation, only whether headings are provided. The large number of both topsites and government sites using headings in the more recent years is no doubt due, in part, to the fact that they are both easy for developers to understand and to code. We note that pages in our full corpus have a surprisingly deep headings structure. For topsites, 8.5% of the headings are at level 1, 40.9% are at level 2, and 50.6% are at level 3. Government sites were similar with 9.8% of the headings at level 1, 42.0% at level 2, and 48.2% at level 3.

Table XI shows the percentage of topsites and government sites in the trends corpus using headings to structure page content. The difference in the percentage of topsites and government sites using headings was not statistically different, $t(26) = .701$, $p > .05$, two-tailed, and for both there were significant increases over the years in the percentage of sites using headings, as reflected in significant positive correlations for both types of sites.

The final method of meeting the criterion for bypassing blocks of material is through the use of (appropriately titled) frames. The number of frames in our full corpus was very small, however, with only 3.5% of all topsites and 1.3% of all government sites using a frameset. For those few sites, there is possibly a greater attention to frame titles for the government sites with all 4 frames in the corpus being titled. This contrasts with the topsites in which only 6 of 53 frames were titled. In any event, frame use is too infrequent to be seen as an alternative to skip navigation and headings in the sites we examined.

4.2.2. Navigable: Page Titles. WCAG 2.0 success criterion 2.4.2 states that all pages should have a title that describes the page's topic or purpose. Such titles identify the

Table XI. Percentage of Sites Using Headings

Year	Unique Topsites	Government Sites
1999	3.0%	20.0%
2000	3.0%	20.0%
2001	0.0%	20.0%
2002	0.0%	10.0%
2003	3.0%	20.0%
2004	21.2%	20.0%
2005	30.3%	40.0%
2006	39.4%	50.0%
2007	54.5%	60.0%
2008	63.6%	80.0%
2009	72.7%	80.0%
2010	78.8%	100.0%
2011	90.9%	100.0%
2012	87.9%	100.0%
Mean	39.3%	48.9%
SE	9.43	9.83
rho	.946**	.836**

Reported in the bottom row are correlations with year, all n = 14, ** $p < .005$, two-tailed.

current location for users and typically appear in the title area of the browser or in a browser tab and can appear in site maps or lists of search results to help identify content. Having page titles was a priority-2 checkpoint in WCAG 1.0. It is a level-A success criterion for WCAG 2.0. Page titles should be of the following form.

```
<title>University Admissions Process</title>
```

The content inside this title element should be useful, not contain unhelpful information such as the name of the file (e.g., "page.html") or place holders supplied by authoring tools (such as "Enter the title of your HTML document here" or "Untitled Document").

Results of our automated check showed that 100% of government pages had a title. For the topsites, only one page was missing a title in 2001, two were missing a title in 2003, and one was missing in 2012.

Human inspection of the titles indicated that none had the specific errors such as "page.html" or "Untitled Document". However, we note that there were some cases where the title clearly contained errors. For example, the title for one site contained characters such as "Ãƒ Âƒ Ã‚ Â¢Ãƒ Â‚ Ã‚ Â€Ãƒ Â‚ Ã‚ Â". For the topsites, there was one such error in 2004 and one in 2012; there were no such errors on government websites. There were instances in which the usefulness or relevance of the title could be questioned. We noted, for example, cases in which the title was "Homepage," "Sign in", and "it.mail" (this title for a newspaper site). Such cases were relatively rare however, with only one for a government site (2004) and from 3 to 7 on topsites (the exact number depending on human judgment about the relevance of the title).

4.3. Understandable

The WCAG 2.0 principle for understandability is designed to ensure that page content and interface controls are understandable to all users. The different aspects of this address issues that affect page readability, the predictability of page appearance and operation, and issues that affect users' ability to avoid and recover from errors. By and

Table XII. Violations of the HTML Language Attribute

Year	Unique Topsites	Government Sites
1999	100.0%	100%
2000	100.0%	100.0%
2001	97.0%	90.0%
2002	97.0%	70.0%
2003	93.9%	70.0%
2004	81.8%	80.0%
2005	75.8%	60.0%
2006	72.7%	60.0%
2007	63.6%	60.0%
2008	60.6%	50.0%
2009	51.5%	40.0%
2010	54.5%	40.0%
2011	45.5%	30.0%
2012	42.4%	10.0%

Shown are the percentage violations for the topsites and government sites, by year.

large, decisions about such factors require human testing. Our machine analysis was therefore limited to only one measure related to readability.

4.3.1. Readability: Language of Page Identified. There is only one level-A success criterion for making a page readable, which is to specify the language attribute of an HTML document, such as

```
<html lang="en">.
```

Identifying this attribute is important, for example, for speech synthesizers that support multiple languages so that they are able to properly pronounce words using the appropriate language features. This was a priority-1 checkpoint in WCAG 1.0 and is a level-A success criterion for WCAG 2.0 (success criterion 3.1.1).

There was no significant difference between the percentage of pages containing the lang attribute between topsites and government sites either for the full corpus [topsites (mean = 28.56%, SE = 6.22) and government sites (mean = 47.81%, SE = 8.04), $t(26) = 1.89$, $p>.05$, two-tailed, $d = .74$] or the trend corpus [topsites (mean = 74.0%, SE = 5.67) and government sites (mean = 61.4%, SE = 7.02), $t(26) = 1.395$, $p > .05$, two-tailed, $d = .55$].

Despite the seeming ease of implementation of these guidelines (it requires only a single line of code per page), there were a very large number of violations on this SC. In 1999 there was no site in the trends corpus (topsites or government sites) that included this language attribute. As shown in Table XII, both topsites and government sites improved over the years, with government sites, in particular, decreasing their violation over the years. For the trend sites, there was a decrease in omitting the lang attribute over the 14 years studied. The percentage of errors for topsites and government sites is shown in Table XII. The decrease in violations was confirmed by significant correlations for both topsites, rho = $-.993$ n = 14 $p < .001$, and government sites, rho = -9.79 n = 14 $p < .001$.

4.4. Robustness

Robustness refers to the fact that the HTML/XHTML code must be robust enough that assistive technologies (user agents) can reliably interpret it. If the code is malformed, user agents will find it difficult to accurately render pages.

Table XIII. HTML/XHTML Parsing Errors

	Unique Topsites			Government Sites		
Year	**# lines of code**	**% issues all types**	**% issues WCAG**	**# lines of code**	**% issues all types**	**% issues WCAG**
1999	10378	25.4%	15.7%	1918	26.3%	18.3%
2000	18931	18.4%	10.0%	2968	24.6%	19.5%
2001	21373	20.5%	10.5%	2965	22.9%	16.1%
2002	18456	29.2%	13.6%	4821	13.9%	9.2%
2003	17845	34.1%	14.6%	10682	8.5%	4.5%
2004	18866	31.3%	10.6%	16136	6.8%	3.1%
2005	20054	27.1%	9.3%	8122	11.7%	5.5%
2006	23233	29.0%	8.0%	7527	12.0%	6.0%
2007	26649	29.3%	7.1%	7754	13.0%	7.0%
2008	29868	24.8%	4.0%	6294	11.2%	7.8%
2009	38892	21.0%	3.6%	7097	8.8%	6.2%
2010	44989	21.2%	3.3%	7566	7.9%	3.4%
2011	41573	21.8%	4.1%	8926	7.1%	3.2%
2012	47618	20.8%	2.3%	15476	8.4%	6.8%
Mean		25.3%	8.2%		13.1%	8.2%
SE		1.28	1.19		1.78	1.48
rho		−.152	−.889**		−.723*	−.552*

Shown are the total number of lines of code per year along with the percentage errors (total and WCAG errors) for the topsites and government sites, by year. Reported in the bottom row are correlations with year, all n = 14, * $p < .05$, ** $p < .005$, two-tailed.

4.4.1. Parsing. We considered the robustness of the Web pages in our corpus by testing for the level-A success criterion of parsing, SC 4.1.1. This was a priority-2 checkpoint in WCAG 1.0. We examined how the websites performed on this criterion in the trend corpus using the HTML Tidy software to list all errors and warnings.

There were significantly more HTML errors for topsites than government sites. Table XIII shows the proportion of total lines of HTML having warnings or errors on topsites and government sites in this corpus, $t(26) = 5.57$, $p<.001$, $d = 2.18$. WCAG 2.0 identifies several kinds of issues that are important with respect to user agent rendering. These include the requirement that elements have complete start and end tags, that elements are properly nested, that they do not contain duplicate attributes, and that IDs are unique. When only failures to meet these requirements are considered, the percentage of errors for the two types of sites did not differ, $t(26) = -.006$, $p>.05$.

In considering the trends, Table XIII shows that the issues as a proportion of total lines of HTML tended to decrease over the years of studies. These decreases were statistically significant in all cases except for all errors types on topsites.

4.5. Accessibility and Good Coding Practices

The analyses presented so far have indicated a mixed picture of website accessibility. On some cases, there is poor adherence to WCAG 2.0 success criteria, while in other cases it is considerably better, particularly in recent years. This inconsistency caused us to wonder about the underlying reasons for these differences. One of the often cited reasons for low adherence to guidelines is a lack of developer awareness of the standards [Kelly et al. 2005; Loiacono et al. 2009]. The low error rates in some cases, however, would seem to argue against lack of awareness. In looking at the specific cases where there were few violations, however, an alternative hypothesis emerged. Specifically, could the instances of accessibility implementation be, in essence, a byproduct

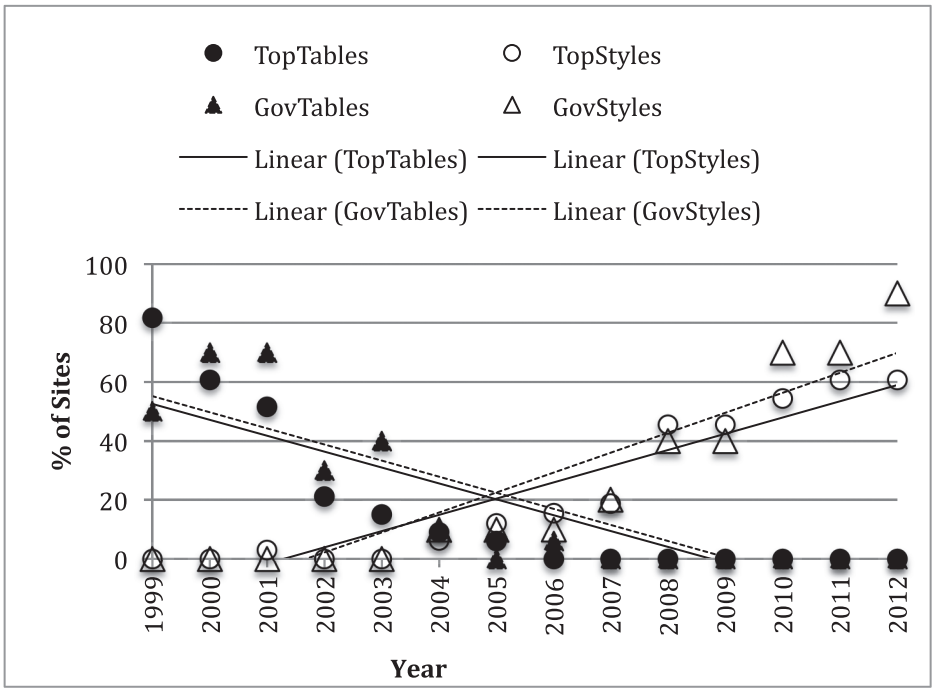

Fig. 4. The percentage of sites using tables and styles for topsites and government sites. Also shown are the linear trend lines for each. Note that the percentage of topsites and government sites does not sum to 100% each year and this figure does not include the percentage of sites that had mixed tables and styles.

of good coding practices and exploitation of newer page design features rather than improved awareness and provision for accessibility?

To consider this question, we looked at an indicator of good coding practice and how it might relate to accessibility. A first clue about changing coding practices comes from the data of the image analysis shown in Table V. The analysis of percentage violations for alternative text appears to have only told part of the story. Looking at the number of images reveals a more complex story. There is a clear decline over the years in the use of decorative images, beginning about 2004 to 2006. As a main use of such images is for controlling the placement of elements on a page, we considered this and other methods of page layout, tables, and style sheets, and how they might have changed over the years.

In the case of our corpus, consisting exclusively of homepages, tables were clearly used to control page layout rather than present tabular data. Figure 4 shows the percentage of sites using tables (but no style sheets) and the percentage of sites using style sheets (but no tables) over the 14 years examined. There is a clear pattern of decreasing table use and increasing style sheet use with no sites using tables exclusively from 2007 onward. The use of styles, either blocked or linked, and either alone or in combination with tables, rose from 12% in 1999 (the year following the publication of RFC 2318 [W3C 1998] defining the text/css media type) to 100% for topsites in 2012, and from 30% in 1999 to 100% for government sites in 2012. Significant negative correlations were found between years and table use for both topsites, rho = $-.936$, n = 14, $p < .005$, and government sites, rho = $-.898$, n = 14, $p < .005$. In contrast, there were significant positive correlations for style sheet use over the years for both topsites, rho = .964, n = 14, $p < .005$, and government sites, rho = .971, n = 14, $p < .005$.

While the use of styles does not guarantee an accessible site, it does avoid some common accessibility problems such as those associated with decorative image alt tags and problematic vocalization of elements within layout tables. While it is possible to use styles in a way that harms accessibility (e.g., using text styles instead of heading tags, or adding a cluttered background image that reduces visual contrast), the separation of document structure from presentation will generally lead to more accessible pages and ones that are also easier for the end user to modify through accessibility transformations (especially user style sheets).

This consideration of methods of page layout suggests that some improvement over the years with respect to accessibility criteria may be related to changes in coding practices.

5. DISCUSSION

It has been reported that disabled individuals often return to the relatively few sites that are known by them to be accessible [Bigham et al. 2007; Takagi et al. 2008]. Given the findings here and elsewhere of the often-low adherence to accessible Web criteria, such strategies may be understandable. As highlighted by Figure 3, adherence to accessibility criteria is low, even for the use of alternative text for images. Alternative text was a level-1 checkpoint for WCAG 1.0 and a level-A success criteria for WCAG 2.0. Moreover, compared to other Web accessibility features, this indicator has received the preponderance of attention, probably due to its testability, its importance for visually impaired users, and the fact that it is relatively easily understood.

The high level of violations we found for alternative text can be contrasted with the findings of Loiacono et al. [2009] in which 27% of the Fortune 100 sites tested were free of priority-1 violations. Figure 3 shows that the number of sites being free of violations on just one criterion is quite low. Our findings also show more violations on individual images than the study of Hackett et al. [2004]. We can compare our findings from Table V with the comparable years of that study. Using their number of errors for alternative text on images, they report 49.7%, 53.0%, 46.1%, 63.0%, 53.9%, and 41.7 % violations over the years from 1997 to 2002.

Why do the current results suggest more violations for alternative text on images than these previous studies? The answer is likely to be found in the algorithms used to measure violations. The current study used quite strict criteria as to whether alternative text was correct. Following Bigham et al. [2006], we divided images, based on likely role, into either content or decorative images. This was not done in the previous studies mentioned. As a result we were able to be more precise as to whether the alternative text on pages was appropriate, not just whether it was present. Our findings for this measure are in line with those reported by Bigham et al. who reported violations for content images on 79.2% of their high-traffic sites.

From our comparison with these earlier studies, we suggest that deeper analysis of accessibility indicators, such as that done in the present study, may reveal more accessibility problems on pages than has been previously indicated. This certainly is not good news. It does, however, suggest that it may be worth considering whether there are additional aspects of accessibility checking that should be added to existing tools.

Despite the fact that there continue to be a number of problems with the accessibility of websites, our relatively long-term study did demonstrate that on all measures examined there was at least some improvement for topsites and often quite striking improvement for government sites. This is summarized in Table XIV. Over the years there are generally fewer violations in the provision of alternative text for images,

image buttons, and input for forms. User agent operability and rendering have improved due to an increased use of skip navigation and headings, increased use of the language attribute, and a reduction in HTML errors. Finally, a growing awareness of accessibility is suggested by an increase in the number of statements regarding site accessibility.

Table XIV also shows that, in general, government sites had fewer accessibility problems than did topsites. This finding is consistent with other work using different sites and different indicators of accessibility [Hackett et al. 2004]. Given the mandate of government sites to serve a broad spectrum of citizens, it is encouraging to find this result.

Signs that the success criteria are incompletely understood, however, come from a number of sources in the current data, most clearly from the analysis of image alt tag errors. For content images, there was evidence, significant for the government sites, that missing alt tags were a decreasing source of errors. Increasingly, alt tag errors on images were the result of empty (not missing) alt tags. For decorative images, errors increasingly were of the form of nonempty alt tag text. While a more subtle error than an empty alt tag string on a content image, this again suggests a failure to understand how screen readers behave. The reasons for these patterns for content and decorative images are not clear, although the pattern suggests some attention to accessibility (either directly by content authors or indirectly by automatic tool insertion), without full understanding of how or why such attributes are applied and what constitutes appropriate text for the tags.

As mentioned in the discussion of WCAG 2.0 success criteria, conformance to these guidelines requires that pages fully meet the success criteria for a level to be compliant at that level. In this study, we have examined only level-A criteria, and only a subset of those criteria: those that are machine testable and relatively easy to satisfy. Even with just these success criteria we find that almost no websites, including government ones, would be considered accessible. Figure 3 shows how few websites were accessible even on just the one indicator of providing appropriate alternative text for images. If testing were broadened to include all level-A success criteria, including those that must be tested by humans, we would imagine that the number of sites considered accessible would be fewer still. The WCAG 2.0 guidelines were designed so that level-A would always being required [Reid and Snow-Weaver 2008], levels AA and AAA compliance being required to meet the needs of all users. Similarly, poorer compliance for priority checkpoints 2 and 3 than priority-1 has been found in other work [Loiacono et al. 2009]. Expanding consideration to full compliance with level-AA and level-AAA criteria it is likely that few, if any, of the sites we tested would be considered compliant.

One especially intriguing finding emerged in the course of this research with respect to the relationship between accessibility and good coding practices. In considering the overall results, there are some cases in which there were few violations of accessibility guidelines: image buttons (especially for the government sites), headings, page titles, and having the language of pages identified (again, especially good for the government websites). These latter two, in particular, are very straightforward and perhaps that is part of the key to their correct use. Given the complexity of some of the success criteria, this may indeed be a factor.

The relatively low number of violations for these specific features, however, made us consider whether there might be another explanation. Having page titles is important for having the page be identifiable in a collection of browser tabs or a set of browser windows and plays a role in being found during a search. Headings, although commonly used to support screen reader navigation [Borodin et al. 2010] are primarily used to control *visual* page rendering. As such it is perhaps not surprising that their

Table XIV. Summary Table of Trends Over the 14 Years Studied

Principle	Success Criteria	Summary of Results
Perceivable	1.1.1 Text alternatives for non-text content	*Images* Few sites were free of violations on providing alternative text, although there was some improvement over the years studied. For individual images, there were decreasing violations for both content and decorative images; overall there were fewer violations for content images than decorative images. For both content and decorative images, missing alt tags decreased over the years, with violations increasingly being due to empty tags (content images) or non-empty tags (decorative images). *Image buttons* For topsites, there was a tendency for the percentage of violations to decrease over the years studied. There were few violations for government sites. *Input form labels* Violations decreased over the years studied for both topsites and government sites. There were fewer violations on government sites.
Operable	2.4.1 Provide means for bypassing blocks of content	*Skip navigation* Skip navigation links were increasingly provided on both topsites and government sites; they were more likely to be used on government sites. *Headings* Headings have been growing in their frequency of use for both topsites and government sites, with the majority of both having provided headings for the past several years. *Frames* In our corpus frames were too infrequent for detailed analysis.
	2.4.2 Provide page titles	100% of the government pages and all but a handful of the topsites had appropriate page titles.
Understandable	3.1.1 Specify language of the page using the HTML language attribute	Strong trends were apparent over the years for both topsites and government sites to decrease violations in providing language information.
	3.3.2 Labels are provided for interactive elements	See 1.1.1 above for input form labels.
Robust	4.1.1 HTML/XHTML parsing errors are avoided	HTML errors decreased over the years studied, particularly on those elements that most affect accessibility. Topsites and government sites did not differ significantly in their number of violations on these elements.
[other]	[Accessibility statements]	Linked accessibility statements and mentions of accessibility on home pages were more common for government than topsites, with government sites, in particular, increasingly having such information on their sites.
	[Style sheets for layout]	Style sheets for were increasingly used for layout by both topsites and government sites. Tables were used less frequently over the 14 years studied.

use by developers appears to be more widespread than the use of skip navigation which is specific to accessibility.

Given these nonaccessibility reasons for having a title and headings, we wondered, first, if success is primarily a side-effect and, second, whether future enhancements in Web technologies could be designed such that accessibility as a side-effect is enhanced. Additional clues come from the analysis of decorative images, tables, and styles as methods of providing page layouts. Specifically, we noted that the use of decorative images and tables decreased over the years, while the use of styles increased dramatically (see Figure 4). Decorative image alt tag errors will, of course, decrease as the use of decorative images decreases. And the accessibility benefits of using styles (in general, and especially as an alternative to tables for layout) will generally accrue whether their use is motivated by accessibility concerns or, as is more likely, a desire to have more control over page design.

These findings from this large corpus support our suggestions from other work. Previously, we reported a detailed manual inspection of four representative government, news, and commercial Websites [Richards et al. 2012]. In that study, accessibility improvements were found to arise from several changes in coding practices. Specifically, these pages exhibited improved accessibility as a side-effect of the use of new browser features to enhance page layout and design, page coding for optimization of page rank in search results, and changes in coding to attain both better cross-browser consistency and cross-device compatibility.

As with all research, the specific methodologies employed limit the conclusions that can be drawn. In the present study, not all success criteria, and not all parts of multipart success criteria (such as with SC 1.1.1), could be machine tested. We also note that our algorithms may differ from those used by others (and are, in some cases, more stringent than automatic accessibility testers). We point, in particular, to the definitions of "content" and "decorative" images that we employed. The success criteria do not state specific image sizes for these to guide automated tools in deciding how to categorize these. We adopted what we believe to be somewhat conservative criteria, not wanting to incorrectly categorize one or the other type of image. In the interests of replicability, all the sites we tested are listed in the Appendix.

The question of whether a page is "accessible" for an individual user is complex and based on a number of factors, such as the user's particular needs and what assistive technology, if any, he/she might be using. One of the factors not considered in the WCAG guidelines is whether a page is actually usable. There is an assumption that to be usable it must be accessible. The inverse is clearly not the case, however, and it is quite possible for pages to be accessible but remain unusable for disabled individuals [Kelly et al. 2005; Leporini and Paternò 2004; Power et al. 2011], and not all ways of satisfying success criteria may be equally useful to disabled individuals [Power et al. 2011]. Thus, while accessibility is a step towards inclusion, much more work is needed if the Web is to be a level playing field for disabled users.

A number of criticisms have been leveled at the WCAG guidelines, both 1.0 and 2.0. The most frequent criticisms focus on the complexity and testability of the guidelines. While the guidelines are very thorough, this thoroughness makes them difficult for developers who are not accessibility experts to understand and thus implement. While WCAG 2.0 was designed to make the guidelines more understandable and testable, recent studies continue to show that nonexperts and even experienced Web accessibility experts have trouble agreeing on their use [Brajnik et al. 2012; Lazar et al. 2011]. The success criteria we tested represent those for which experienced accessibility auditors agree on conformance [Brajnik et al. 2012]. Specifically, titled pages (SC 2.4.2), providing alternatives for nontext (SC 1.1.1), and identifying the language of pages (SC 3.1.1) have high agreement. There may be less agreement about how

to use bypass blocks effectively (SC 2.4.1), but the marker we used is at least easily testable and its absence certainly means that the page is not as accessible as it could be.

Finally, we return to the fact that the pages tested in this study were only the homepages of websites. As such, this does not indicate how well the full sites might fare on the accessibility indicators examined. It is likely, however, that this work gives a more optimistic view of site accessibility than is actually the case. As mentioned, homepages are generally the ones most honed and most likely to be accessible [Nielsen 2000].

6. CONCLUSIONS

Our findings are consistent with previous results for the years studied in common; both topsites and government sites exhibit generally low conformance with Web accessibility indicators. This is especially true in the years prior to the introduction of WCAG 2.0, the period examined in previous studies. Looking over the entire 14 years since WCAG 1.0, however, we see evidence of improvement, at least for some success criteria, with government sites showing more improvement than topsites on some, but not all, criteria.

A more detailed examination of both successes and failures leads us to question the extent to which accessibility gains can be attributed to awareness of and adherence to WCAG guidelines, however. At least some improvements appear to be side-effects of good coding practices and the desire to improve Web page design and website prominence in search results. Future improvements in accessibility may come from explicitly designing new technologies to have such beneficial side-effects.

APPENDIX

Following is a complete corpus of Websites used in this research. The topsites are ordered by decreasing traffic volume as of April 19, 2011 [Alexa 2011].

Topsites			
UK websites	**Missing Years**	**US websites**	**Missing Years**
google.co.uk	2005 is 12/31/04	google.com	
facebook.com	1999–2010[1]	facebook.com	1999–2010[1]
google.com		yahoo.com	
youtube.com	1999–2004	youtube.com	1999–2004
bbc.co.uk		amazon.com	2006 is 12/27/05
yahoo.com		wikipedia.org	1999–2000
ebay.co.uk		twitter.com	1999–2005
live.com	1999–2005	blogspot.com	1999–2010[2]
wikipedia.org	1999–2000	ebay.com	
twitter.com	1999–2005	craigslist.org	
amazon.co.uk	1999	live.com	1999–2005
blogspot.com	1999–2010[2]	linkedin.com	1999–2003, 2004, 2006
linkedin.com	1999–2003		is 10/26/05
msn.com		msn.com	1999
paypal.com		bing.com	1999–2009
dailymail.co.uk	2007 is 12/27/06	go.com	
wordpress.com	1999–2004	cnn.com	1999, 2006 is 12/31/05
guardian.co.uk		aol.com	
bing.com	1999–2009	wordpress.com	1999–2004, 2008 is
flickr.com	1999–2003		12/30/07
imdb.com	1999	espn.go.com	
hsbc.co.uk		paypal.com	
amazon.com		netflix.com	2001, 2002, 2007, 2000
apple.com			is 11/28/99
microsoft.com		weather.com	2002 is 12/17/01
telegraph.co.uk	1999–2010[2]	flickr.com	1999–2003

rightmove.co.uk	1999	imdb.com	1999
gumtree.com	1999–2004, 2011 is from 02/08, 2012 is from 04/23	nytimes.com microsoft.com	
goggleusercontent.com	1999–2010[2]	apple.com	
groupon.co.uk	1999–2008	bankofamerica.com	
sky.com	1999–2000, 2002 is 12/17/01 2008 is 11/18/2007	godaddy.com tumblr.com	1999, 2002 1999–2006
lloydstsb.com	1999	about.com	
virginmedia.com	1999–2000, 2002	chase.com	1999
barclays.co.uk		huffingtonpost.com	1999–2005[6]
tumblr.com	1999–2006	optmd.com	1999–2008
national–lottery.co.uk	2010	zedo.com	1999
livejasmin.com	1999–2001	ehow.com	
direct.gov.uk	1999, 2001, 2002, 2009, 2008 is 12/18/07	securesever.net ask.com	1999–2010[2]
wordpress.org	1999–2002	goggleusercontent.com	1999–2010[2]
natwest.com	1999–2006, 2010	yelp.com	1999–2003, 2005 is 10/28/04
lloydstsb.co.uk	1999–2012[3]		
aweber.com		myspace.com	
tesco.com	1999–2003	foxnews.com	
about.com		comcast.net	2005–2008
xhamster.com	1999–2006	pornhub.com	1999–2000, 2002–2006
pornhub.com	1999–2000, 2002–2006	aweber.com	1999–2008
autotrader.co.uk	1999–2001	imgur.com	
thesun.co.uk	1999–2010[1]	hulu.com[7]	
argos.co.uk	1999–2003, 2010	mozilla.com[4]	
mozilla.com[4]		wellsfargo.com	2007 is 12/18/06, 2008 is 12/27/07
pgmediaserve.com	1999–2012[5]		
bt.com		reddit.com	1999–2004
partycasino.com	1999–2002, 2004	wordpress.org	1999–2002
aol.com		cnet.com	
moneysavingexpert.com	1999–2002	answers.com	1999–2003
nwolb.com	1999–2004, 2009–2010	etsy.com	1999–2003
thepiratebay.org	1999–2003	livejasmin.com	1999–2001
ebay.com		references.com	2010
adobe.com		xhamster.com	1999–2006
skysports.com	1999–2000, 2005 is 8/26/04, 2007 is 11/17/06	adobe.com usps.com walmart.com	
Government sites			
dwp.gov.uk	1999–2000	ssa.gov	
fco.gov.uk		fema.gov	
direct.gov.uk[8]	1999–2003	medicare.gov	
dhs.gov.uk	1999–2001, 2004–2006	ed.gov	
education.gov.uk	1999–2009	fta.dot.gov	
environment-agency.gov.uk		justice.gov whitehouse.gov	1999–2000
justice.gov.uk	1999–2006	house.gov	
parliament.uk		irs.gov	1999
hmrc.gov.uk	1999–2004	dhs.gov	1999–2001, 2004–2006
homeoffice.gov.uk	2001, 2003		

[1] nothing in Web Archive due to robots.txt, 2011 and 2012 from manual sample only
[2] nothing usable in Web Archive, 2011 and 2012 from manual capture only
[3] not used since content duplicated lloydstsb.com
[4] mozilla.org archived for all years
[5] nothing in Web Archive and no current site
[6] 2005 site was "coming soon"
[7] major site transition in 2008 (probable change in domain name ownership)
[8] both a UK topsite and government site

ACKNOWLEDGMENTS

The authors would like to thank Dr. Lorna Gibson for providing input during the initial Website selection and the SiDE project team for discussions about this work. We thank the reviewers for their helpful input and Kevin Carey for interesting insights related to issues raised in this article.

REFERENCES

ACM Curricula Recommendations. 2012. http://www.acm.org/education/curricula-recommendations

Alexa. 2011. http://www.alexa.com/topsites/countries

Alonso, F., Fuertes, J. L., González, and Martínez, L. 2010. On the testability of WCAG 2.0 for beginners. In *Proceedings of the International Cross Disciplinary Conference on Web Accessibility (W4A'10)*. ACM, Article 9.

Bigham, J. P., Cavender, A. A., Brudvik, J. T., Wobbrock, J. O., and Lander, R. E. 2007. WebinSitu: A comparative analysis of blind and sighted browsing behavior. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'07)*. ACM, 51–58.

Bigham, J. P. Kaminsky, R. S., and Ladner, R. E. 2006. WebInSight: Making web images accessible. In *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'06)*. ACM, 181–188.

Brajnik, G., Yesilada, Y., and Harper, S. 2012. Is accessibility conformance an elusive property? A study of validity and reliability of WCAG 2.0. *ACM Trans. Access. Comput. 4*, 1, Article 5.

Borodin, Y., Bigham, J. P., Dausch, G., and Ramakrishnan, I. V. 2010. More than meets the eye: A survey of screen-reader browsing strategies. In *Proceedings of the International Cross Disciplinary Conference on Web Accessibility (W4A'10)*. ACM, Article 13.

Chen, A. Q. 2009. Web evolution. *SIGWEB Newslett.* Summer, Article 4.

Davis, J. J. 2002. Disenfranchising the disabled: The inaccessibility of internet-based health information. *J. Health Comm. 7*, 4, 355–367.

Disability Rights Commission. 2004. *The Web: Access and Inclusion for Disabled People.* The Stationery Office, London.

Ellison, J. 2004. Assessing the accessibility of fifty US government web pages – Using bobby to check on uncle sam. *First Monday 9*, 7. http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1161/1081.

Forrester Research, Inc. 2003. The wide range of abilities and its impact on computer technology. http://www.microsoft.com/enable/download/default.aspx#research.

Hanson, V. L., Richards, J. T., Harper, S., and Trewin, S. 2009. Web accessibility. In *The Universal Access Handbook*, C. Stephanidis Ed., CRC Press, Boca Raton, FL, 10-1–10-15.

Harper, S. 2008. Web evolution and its importance for supporting research arguments in web accessibility. In *Proceedings of the 1st International Workshop on Understanding Web Evolution (WebEvolve'08)*. http://journal.webscience.org/40/.

Hackett, S., Parmanto, G., and Zeng, X. 2003. Accessibility of internet websites through time. In *Proceedings of the 6th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'04)*. ACM, 32–39.

Hong, S., Katerattanakul, P., Choi, H. R., Kang, Y. S., and Cho, J. 2007. Evaluating government website accessibility: A comparative study. In *Proceedings of the International DSI and Asia and Pacific DSI Conference.*

Kelly, B., Sloan, D., Phipps, L., Petrie, H., and Hamilton, F. 2005. Forcing standardization or accommodating diversity? A framework for applying the WCAG in the real world. In *Proceedings of the International Cross-Disciplinary Workshop on Web Accessibility (W4A'05)*. ACM, 46–54.

Lazar, J., Beere, P., Greenidge, K., and Nagappa, Y. 2003. Web accessibility in the Mid-Atlantic United States: A study of 50 homepages. *Universal Access Inf. Soc.*, 331–341.

Lazar, J., Wentz, B., Bogdan, M., Clowney, E., Davis, M., Guiffo, J., Gunnarsson, D., Hanks, D., Harris, J., Holt, B., Kitchin, M., Motayne, M., Nzokou, R, Sedaghat, L., and Stern, K. 2011. Potential pricing discrimination due to inaccessible web sites, Part 1. In *Proceedings of the INTERACT Conference*. P. Campos et al. Eds., Lecture Notes in Computer Science, vol. 6946, Springer, 108–114.

Leporini, B. and Paternoò, F. 2004. Increasing usability when interacting through screen readers. *Universal Access Inf. Soc. 3*, 57–70.

Loiacono, E. T., Romano, N. C. Jr., and McCoy, S. 2009. The state of corporate website accessibility. *Comm. ACM 52*, 9, 128–132.

Ludi, S. 2007. Introducing accessibility requirements through external stakeholder utilization in an undergraduate requirements engneering course. In *Proceedings of the 29th International Conference on Software Engineering.* IEEE Computer Society, 736–743.

Marincu, C. and McMullin, B. 2004. A comparative assessment of web accessibility and technical standards conformance in four EU states. *First Monday 9*, 7, 5. http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1160/1080.

Nielsen, J. 2000. *Designing Web Usability.* New Riders Publishing, Indianapolis, IN.

Petrie, H., King, N., Hamilton, F., and Weisen, M 2005. The accessibility of museum websites: Results from an English investigation and international comparison. In *Proceedings of HCI International.* Vol. 8. *Universal Access in HCI: Exploring New Dimensions of Diversity*. Mira Digital Publishing.

Poor, G. M., Leventhal, L. M., Barnes, J., Hutchings, D. R., Albee, P., and Campbell, L. 2012. No user left behind: Including accessibility in student projects and the impact on CS students' attitudes. *ACM Trans. Comput. Educ. 12*, 2, Article 5.

Power, C., Petrie, H., Freire, A. P., and Swallow, D. 2011. Remote evaluation of WCAG 2.0 techniques by web users with visual disabilities. In *Universal Access in HCI*, Part I, *HCII*, C. Stephanidis Ed., Lecture Notes in Computer Science, vol. 6765, Springer, 285–294.

Reid, L. G. and Snow-Weaver, A. 2008. WCAG 2.0: A web accessibility standard for the evolving web. In *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility (W4A'08)*. ACM, 109–115.

Richards, J. T., Montague, K., and Hanson, V. L. 2012. Web accessibility as a side effect. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'12)*. ACM.

Sampson-Wild, G. 2007. Testability costs too much. A list apart. http://www.alistapart.com/articles/testability/.

Takagi, H., Kawanaka, S., Kobayashi, M., Itoh, T., and Asakawa, C. 2008. Social accessibility: Achieving accessibility through collaborative metadata authoring. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'08)*. ACM, 193–200.

W3C. 1998. The text/css media type. http://tools.ietf.org/html/rfc2318.

W3C. 2006a. Complete list of web accessibility evaluation tools. http://www.w3.org/WAI/ER/tools/complete.

W3C. 2006b. Policies relating to web accessibility. http://www.w3.org/WAI/Policy/.

W3C. 2010a. Developing a web accessibility business care for your organization: Overview. http://www.w3.org/WAI/bcase/Overview.html/.

W3C. 2010b. Web accessibility and older people: Meeting the needs of ageing web users. http://www.w3.org/WAI/older-users/.

W3C. 2011a. Web accessibility initiative (WAI). http://www.w3.org/WAI/.

W3C. 2011b. Web content accessibility guidelines (WCAG) overview. http://www.w3.org/WAI/intro/wcag.php.

Waller, A., Hanson, V. L., and Sloan, D. 2009. Including accessibility within and beyond undergraduate computing courses. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'09)*. ACM, 155–162.

Wayback Machine. 2012. http://www.archive.org/web/web.php.

World Health Organization (WHO). 2013. http://www.who.int/mediacentre/factsheets/fs352/en/index.html.

Zeng, X. and Parmanto, B. 2004. Web content accessibility of consumer health information web sites for people with disabilities: A cross sectional evaluation. *J. Med. Internet Res. 6*, 2, e19.

Received March 2012; revised September 2012; accepted November 2012