

# Comparing and evaluating the effectiveness of mobile Web adequacy evaluation tools

Arvinder Kaur<sup>1</sup> · Diksha Dani<sup>2</sup>

Published online: 2 May 2016  
© Springer-Verlag Berlin Heidelberg 2016

**Abstract** The number of Web sites is growing exponentially and so are the people who are accessing them on mobile devices including people with special abilities. Mobile-friendly Web sites further increase traffic to the Web site, customer participation, and boost e-Commerce. In such a scenario, it is important to ensure that the Web sites are accessible on mobile devices for all users. Automatic tools provide an easy and a less error-prone mechanism for evaluating the mobile adequacy of these Web sites. The present study focuses primarily on the comparison between four tools of mobile Web adequacy evaluation, namely mobileOK checker, *taw*, EvalAccess Mobile, and mobiReady. The comparison is made based on three parameters, i.e., correctness, completeness, coverage with respect to the conformance of Mobile Web Best Practices (MWBP), and four groups of mobile Web best practice guidelines, i.e., Navigability and Links, Page Layout and Content, Page Definition, and User Input. Intrareliability of the tools and inter-reliability between the tools are also investigated. The improvement required in MWBP and the role of various mobile platforms in mobile adequacy today are also discussed. The study was conducted using automatic tools, manual evaluation, and statistical tools. The results of the study indicate that the coverage of the automated tools was found to lie between 45 and 68 %. The completeness ranged from 14 % (by MobileOK) to 59 % (by EvalAccess mobile). The correctness of the tools

ranged from 42 to 51 %. Statistical analysis highlighted that the tools showed no significant differences in reporting false positives, false negatives and true positives, except *taw* and EvalAccess which reported some difference in reporting true positives. The intrareliability evaluation showed that the tools were highly stable. Intrareliability evaluation done by Krippendorff's alpha revealed that *taw* and mobileOK show the similar results while evaluating the MWBP guidelines in most cases. However, for more accurate evaluations manual evaluation along with automatic evaluation must be performed. The practical significance of such a study can lead to improving the guidelines, modification of the tools, proactive use of the tools, and an overall awareness about building the Web sites accessible for all.

**Keywords** Mobile Web accessibility · Mobile Web accessibility evaluation tools · MWBP · MobileOK

## 1 Introduction

The term “mobile Web” refers to access to Internet services using a mobile device connected through mobile or other wireless networks. The mobile revolution has been an accelerator for many online users. Mobility enables users to access information at any place and any time. The majority of people in the world still cannot afford personal computers, though due to the low cost, the mobile Web has been able to penetrate into almost all the income groups and the remote areas. The decreasing costs of mobile phones and mobile call data rates have likewise contributed to the increased use of mobile devices for internet access. The number of active mobile connections has surpassed the total world population [1]. In fact, mobile Internet usage

---

✉ Diksha Dani  
diksha3108@gmail.com

<sup>1</sup> Guru Gobind Singh Inderprastha University, Sector 16C, Dwarka, New Delhi, Delhi 110078, India

<sup>2</sup> Inderprastha Engineering College, 63- site IV, Ghaziabad, India

has surpassed desktop Internet usage [2]. The use of the mobile Web has grown phenomenally due to the improvement in mobile hardware and software. In fact, most of the new Internet users first used the Internet on mobile phones.

According to the statistics, 91 % of all people on earth have a mobile phone, 56 % own a smart phone, 50 % of mobile phone users use a mobile phone as the primary means to access the Internet and 80 % of the time spent on mobiles is spent using mobile applications [3]. According to 2013 E-Expectations Report of the Higher Education Consultants Noel-Levitz, 68 % of college applicants have viewed college Web sites on mobile devices. Of these students, 24 % would drop a college Web site after a poor Internet experience [4].

Mobile commerce now accounts for 23 % of online sales. Companies with mobile optimized Web sites triple their chance of mobile commerce to 5 % or more. About 43 % of customers are not likely to return to slow loading Web sites; 40 % of the customers are likely to go to a competitor Web site after a bad mobile experience; and 97 % of mobile shopping carts are abandoned due to distraction from unnecessary elements [5]. The reasons due to which shopping carts are abandoned include complex design both visually and functionally, the need to fill long forms, forcing users to sign up, unpredictable checkout processes, and insecurity over payment options [6].

Mobile devices are increasingly becoming an important and primary form of getting information from the Web in many parts of the world. In the present scenario, people use the mobile Web for carrying out banking transactions, making reservations, online shopping, and local searches. The prolific rise in Internet usage offers much easier access to information than in the past, even for people with disabilities. However, people with disabilities still face many difficulties when they try to access the Web on mobile devices. For instance, people with limited mobility may have a hard time controlling a mouse to click on links, and tabbing through menus can be very slow when a cluttered menu on the small screen poses a limitation. Likewise, blind or partially blind people find it difficult to view the screen display. Mobile Web accessibility is a problem not only for people with disabilities, but also for able-bodied people who sometimes face similar problems under certain circumstances such as noise, poor lighting, or glare, while moving, driving, illness, etc.

With the tremendous increase in mobile device use for Internet access, the manufacturers of mobile devices are making an effort to make even the lowest specification device Web enabled and accessible. Therefore, it is extremely important to evaluate the accessibility of Web content on mobile devices.

Web content can be delivered to a mobile device either through a mobile-compatible Web site which can be accessed by using a browser or mobile application which can be downloaded and installed on mobile devices. Since a mobile Web site is instantly accessible to users via a browser across a range of devices, it has a broader reach. Therefore, in the present study tools evaluating the accessibility of Web content delivered on mobile devices are considered.

There are two ways to evaluate the compliance of Web sites with mobile Web accessibility guidelines: manual and automated. Manual evaluation includes inspecting the code and interacting with Web sites by accessibility experts, and sample users, as well as usability testing by participants with or without disabilities. Automated evaluation is carried out by using automated tools.

Nevertheless, manual evaluation of Web sites requires lots of effort, time, and cost. It may as well be subjective with narrow coverage and may diverge with time and expert interpretation. On the other hand, automated evaluation tools provide a quick and cost-effective way of evaluating the mobile Web adequacy with consistent objective evaluation and a broad coverage. However, automated tools can only evaluate the site after it has been implemented though no automated assessment can be applied throughout the Web site design and development process. Moreover, automated tools cannot assess user satisfaction [7]. Vigo et al. also highlight the harm of bypassing human evaluation in favor of evaluation tools, in order to assess the accessibility of a Web site like false positives. The researchers highlight that subjective nature of some guidelines requires expert interpretation of the criteria [8]. It is thus important to assess a Web site using both automated and manual techniques for a more comprehensive evaluation. Therefore, it is also important to assess how effective are these tools in evaluating the accessibility of Web sites on mobile devices.

Mankoff et al. presented and compared various techniques of finding the Web accessibility problems particularly faced by the blind. The researchers compared a laboratory study with blind users to an automated tool, expert review by Web designers with and without a screen reader, and remote testing by blind users. It was found that no single evaluator or tool could be counted on to detect a high percentage of accessibility problems of any type. However, multiple evaluators, working independently, performed better than individuals, particularly using screen readers, which were most consistently successful at finding most classes of problems. It was also found that remote study with blind users was one of the least effective methods [9].

The accelerated growth in the number of Web sites and dynamic rich Web content that occur every day requires frequent updating. Due to a shorter life cycle of Web sites,

the designers and developers of Web sites either do not put enough effort on increasing the mobile Web adequacy or rely heavily on automated tools for evaluating the adequacy of Web content on mobile devices.

There are many tools available to evaluate the mobile Web adequacy of mobile-compatible Web sites. They offer a quick and easy way to evaluate and improve the user experience on the Web for users with or without disabilities when accessed by mobile devices.

This paper aims to investigate how well the tools mobileOK, taw, EvalAccess mobile and mobiReady evaluate the adequacy of the Web sites on mobile devices based on their conformance to MWBP guidelines proposed by the Best Practices Working Group (BPWG). The tools are compared on the basis of three parameters namely, coverage, completeness, and correctness. These parameters have been used in the literature for the evaluation of tools [8].

Though Vigo et al. compared and evaluated the adequacy of Web accessibility tools [8] to the best of the authors' knowledge, there is no study to evaluate the adequacy of tools evaluating the accessibility of Web sites delivered on mobile devices for conformance to Mobile Web Best Practices. Web Content Accessibility Guidelines (WCAG) and MWBP both aim to improve the Web interaction of users who experience barriers due to either disabilities or the device used to access the Web. Although the two sets of guidelines significantly overlap in many areas, there are many differences between the two. While criteria such as non-text alternatives, valid markup, tab order, use of color, and structure are fully covered in WCAG, there are certain guidelines of MWBP which are mobile specific and not related to any WCAG success criteria like Access Keys, Cache Headers, Character Encoding, Input (default input mode), Image maps, Image size, Frames, Page Size, tables, Scrolling among others. These differences are specified in a w3 document [10].

The rest of the paper is organized as follows: Sect. 2 presents an overview of work in mobile Web accessibility, section, Sect. 3 presents an overview of Mobile Web guidelines, Sect. 4 lists out the research questions, Sect. 5 describes the methodology and evaluation tools, Sect. 6 presents the assumptions, Sect. 7 discusses the overall results, Sect. 8 addresses issues with Mobile Web Best practices and their interpretation by the tools, Sect. 9 gives the results of inter and intra reliability tests, Sect. 10 describes the threat to validity, and finally, Sect. 11 concludes the paper.

## 2 Literature review

There are not many studies investigating the adequacy of Web sites on mobile devices. Accessibility of the mobile Web content is an issue faced not only by people with

disabilities, but also by able-bodied people. Yesilada et al. [11] investigated the common barriers between mobile users and the disabled using the barrier walkthrough technique and found that 58 % of the true barrier types were identified as common between the able-bodied mobile users and people with disabilities.

Vigo et al. presented a tool for evaluating the Web adequacy for mobile devices regardless of their software, hardware, or user agent characteristics. The tool was based on mobileOK Basic tests 1.0, which is a subset of Mobile Web Best Practices. The tool was then extended to include device characteristics for the evaluation of Web sites. A case study was then conducted to demonstrate the feasibility of the tool and it was concluded that the tool reduced the number of false positives and false negatives [12].

In another study of mobile Web adequacy evaluation, Bandeira et al. [13] presented a new approach for mobile Web and desktop Web content adequacy evaluation according to the capability and disability of users. They also introduced a tool to evaluate the accessibility for mobile and desktop Web sites according to different disabilities such as blindness, sight damage, deafness, etc., as a validation of the suggested approach. This approach provides support to Web developers and designers to conduct quick specialized accessibility assessments according to different disability types of mobile and desktop environment.

Bandeira et al. also studied the adequacy of Mobile Web technologies to people with disabilities [14] in general and visual impairments in particular [15] discussing the main challenges in attaining the mobile Web adequacy.

Vigo et al. also proposed a tool-supported framework for the assessment of Web sites in terms of conformance with Mobile Web Best Practices 1.0. The researchers also conducted a case study to measure the impact of the assessment framework for desktop and mobile versions of Web sites. The evaluation was carried out using two different mobile devices, one having more features and software support than the Default Delivery Context (DDC) while another having fewer features than the Default Deliver Context. The outcome of the study established that mobile Web pages on more capable devices score higher and that complying to the MWBP increases the usability of Web pages [16]. DDC is defined by Best Practices Working Group as the minimum environment and specifications for the reasonable Web experience even on most basic mobile devices [17].

There are few studies that compare Web adequacy tools. In one such study, Vigo et al. presented the comparison of a pair of tools based on correctness, completeness, and specificity and concluded that though there was room for improvement in the method, the tools were capable of providing accurate and reliable conclusions [18]. Later, in

another study of comparing Web accessibility tools, Vigo et al. [8] empirically evaluated six automatic Web adequacy tools in terms of coverage, completeness, and correctness and concluded that though the use of tools does reduce time and increase the efficiency of identifying the accessibility barriers on a Web site, relying on tools alone and leaving out human judgement were not recommended.

There is a constant rise in the use of mobile devices for carrying out financial transactions, running searches, making travel reservations, online social networking, entertainment, and getting information. There is a need to evaluate how well a Web site would perform on mobile devices. Since the lifecycle of Web site updating is small, automatic tools offer a quick, easy, inexpensive and consistent way for mobile Web evaluation. Therefore, it is important to evaluate how well these tools evaluate the mobile friendliness of the Web sites.

To the best of the authors' knowledge, there are no studies to investigate how well mobile Web adequacy evaluation tools evaluate a Web site for conformance to MWBP. Effectiveness of tools in the present study is measured in terms of coverage, completeness, and correctness. A tool is considered to be effective if it maximizes coverage, completeness, and correctness while minimizing incorrectness. The fruitful implications of this study can lead to improved guidelines, policies, tools, and overall awareness of mobile Web adequacy.

### 3 Mobile Web guidelines

There are no universally accepted standards for evaluating mobile Web adequacy. Existing guidelines such as WCAG 1.0 [19] and WCAG 2.0 [20] have been adapted for delivering Web content on mobile devices. In 2005, the World Wide Web Consortium (W3C) created the Mobile Web Initiative (MWI) which consists of several working groups like the Best Practices Working Group (BPWG) to increase recognition of standards and best practices of publishing on the Mobile Web. With the objective to enhance the user experience of Web sites on mobile devices, the Best Practices Working Group (BPWG) developed and proposed the Mobile Web Best Practices (MWBP). The first working draft of MWBP was prepared in October 2005. After a series of updates on the working draft, the first set of MWBP 1.0 was released in June 2006. The latest version of MWBP 1.0 came out in July 2008 [21]. An extended set of guidelines for MWBP 1.0 was released as a public working group note in 2009 which supplemented MWBP 1.0 by providing additional evaluation of conformance to best practices and a more objective and clear interpretation of Best Practice statements [22].

MWBP is a set of 39 checkpoints or success criteria arranged into five categories. These include the following:

1. *Overall Behavior* These are some general principles for the delivery of Web content on mobile devices, independent of its features.
2. *Navigability and Links* These guidelines define how the structure and the navigation model, including hyperlinking of a Web site, should be done considering the display limitation and input mechanism.
3. *Page Layout and Content* These guidelines focus on the design, the language used in its text, and the spatial relationship between constituent components. It does not address the technical aspects of how the delivered content is constructed.
4. *Page Definition* These guidelines define how page elements such as title, tables, and structural elements, should be defined to exploit the Web technology features for the mobile Web.
5. *User Interface* These guidelines relate to user input in mobile devices and are more restrictive in nature than in a desktop or laptops.

The guidelines also address various issues such as presentation on small screen sizes, input capabilities, and advertising concerns like pop-ups, banners, bandwidth, and cost. The target audience for MWBP includes developers, graphic designers, and mobile Web site maintenance teams [21].

Of these MWBP, the Best Practices Working group extracted two subsets of guidelines – mobileOK Basic [23] and mobileOK Pro test [24]. The mobileOK Basic tests 1.0 are a subset of MWBP guidelines which can be automated and hence, machine verifiable. The mobileOK set of guidelines aims to promote user's Web experience on a variety of mobile devices including a mobile device with the very basic Web capability. MobileOK Basic 1.0 was the second deliverable from the Best Practices Working Group in 2008.

The mobileOK Pro tests include the mobileOK Basic tests and are based on a larger subset of the Mobile Web Best Practices. These tests are not all machine verifiable and require human evaluation. They provide a set of guidelines that fill the gaps left by the limitations of automated tests and thus complete the set of Best Practices. mobileOK Pro set of guidelines is a draft document and work for recommendations is still in progress.

### 4 Research questions

The overall objective of the study is to evaluate the adequacy of mobile Web accessibility evaluation tools. The following questions are investigated during the study:

1. How effective are the tools for evaluating the accessibility of Web sites on mobile devices in terms of coverage, completeness, and correctness?
2. Is there any difference between the true positives, false negatives, and false positives reported by these tools?
3. How relevant are the Mobile Web Best Practices today?
4. What is the level of inter- and intrareliability of the tools?

## 5 Methodology

This study compares mobile Web adequacy evaluation tools. The tools used in this study are W3C's MobileOK checker, EvalAccess mobile, mobiReady, and *taw*. The tools were selected by taking reference from related literature where these tools have been mentioned [18].

The present study was conducted on mobile-compatible Web sites. A Web site is a Web application accessed using a browser. On the contrary, a mobile application is downloaded and resides on the mobile device.

The Web sites selected for the study were taken from *Cantoni.mobi* offers links of high-quality Web sites suitable for mobile devices. The Web sites listed on *Cantoni.mobi* are based on broad usability and interest and are constantly updated. The Web site [www.cantoni.mobi](http://www.cantoni.mobi) lists Web sites in eleven categories, namely business, entertainment, Information, news, portal, search, shopping, sports, technology, travel, and weather. For the present study, one Web site from each category was selected arbitrarily. This was done to include Web sites of distinct domains. From each Web site two pages, a home page and another were selected since according to Hackett et al. [25] home pages is not a true indicator of Web accessibility. The second page of each Web site was selected using a simple random sampling technique. This technique was chosen as each member of the population has an equal chance of being selected independently of the other members of the population [26]. Each of the selected pages were then evaluated by all the tools.

Expert evaluations were carried out by a team of three people, two of whom were working as faculty members and one was from industry. All the experts had more than 5 years of experience. The experts were postgraduates in computer science and engineering. All had experience in Web technologies like HTML, CSS, JavaScript, XML, Ajax in teaching engineering students (as in the case of faculty members), application development (in the case of experts from the industry), evaluation tools and had sound knowledge of MWBP guidelines. They also attended a 1-day workshop on mobile Web accessibility for better understanding of MWBP guidelines.

Experts first evaluated the online Web sites independently to identify the actual violations, according to MWBP conformance. The results were subsequently discussed for the final outcome of actual violations. The team used various techniques to evaluate the Web pages ranging from direct observation by manually reading the source code to using a browser extension (Firefox Accessibility evaluation Toolbar, Chrome), as well as Mobile phone emulators (Mite from Keynote). The violations identified by one tool or technique were cross-verified by other techniques. The automatic evaluation of the Web site by the tools was done immediately after the manual evaluation of the Web page so that there was a minimum time lag between manual and automated evaluation.

### 5.1 Automated tools

MobileOK checker is an online service for performing various tests on a Web Page to determine its level of mobile friendliness based on mobileOK Basic Tests 1.0 specifications. It takes the URL of the Web site as an input and gives the details about the best practices violated, its reason, location, and severity. It also provides suggestions on how the violation can be removed [27]. A Web Page is mobileOK-verified when it passes all the tests.

EvalAccess mobile is an online service to evaluate the conformance to MWBP 1.0 [28]. It takes the URL of the Web site as an input and output are shown in a table with the followings information:

1. Checkpoint or success criteria for which violation has occurred;
2. Description of the checkpoint, name of the HTML attribute that has to appear or has the error/warning;
3. URL of the MWBP guideline in the W3C site where the violated guideline is explained;
4. List of line numbers in the source code where this error/warning has been generated.

*Taw* is an online mobileOK Basic tests 1.0 checker developed by the Centre for the Development of Information and Communication Technologies foundation from Spain (CTIC) [29]. It takes the URL of the Web site as input and lists the instances of each violation type that is identified by the tool and line number in the source code where the violation was detected. It does not give the checkpoint number of the MWBP which has been violated.

MobiReady is an online tool developed by the company dotMobi based on mobileOK 1.0 tests. dotMobi is an expert provider of mobile and Web technology and also involved with the W3C Mobile Web Initiative (MWI) to help formulate the MWI Best Practices for mobile content.

MobiReady is a testing tool that evaluates conformance to mobileOK guidelines and dotMobi compliance. It gives



a mobile readiness score between 1 and 5, five being the best score and one being the worst. It also gives the estimated per kB cost, size, and speed of the Web site in terms of the euro, KB, and kB/second, respectively. It gives the clickable summary of the violations. On clicking the specific violation, it provides the description and instances of the violation. It also generates suggestions on how to fix the violations. However, similar to *taw* it also does not give the checkpoint number of the MWBP that has been violated. It does, however, provide details of checkpoints that are passed. It has two modes of operation, i.e., the page mode, which tests the single page, and the site mode, which tests the entire Web site [30]. For the present study, the page mode *mobiReady* tool has been used as all the other tools under investigation checked one page at a time.

## 6 Assumptions

During the evaluation of the Web sites using the tools specified, the following observations and consequently some assumptions were made for calculating true positives during expert evaluation.

According to the MWBP guideline “4.12 Character Encoding,” all html documents require that the character encoding is specified for each page. There are two ways to specify the character encoding used in a Web page. The HTML5 specification requires that all the meta elements fit in the first 1024 bytes, of the document and the same must be included at the top of the head element as `<meta Charset=“utf-8”>`. The Charset attribute is new in HTML5 and replaces the need for a pragma directive of HTML4, which is specified as `<meta http-equiv=“Content-Type” content=“text/html; charset=UTF-8”>`. In the present study, the violation is considered only if the character encoding is missing; i.e., it is not considered a violation if the encoding is either given in either HTML5 or HTML4 format.

The image size violation is considered to be one violation for the same image if the width or height or both are missing; i.e., height specification and width specification missing are not considered as two separate violations for the same image.

## 7 Overall results

This section investigates the answer to the research question 1, whether the tools are effective in evaluating the adequacy of Web sites on mobile devices in terms of coverage, completeness, and correctness.

As mentioned earlier, the MWBP are categorized into five groups, namely the Overall Behavior, Navigation and Links, Page Layout and Content, Page Definition, and User

Input. Out of these categories, the Overall Behavior checkpoints were not implemented by any of the tools since these are abstract general principles for the delivery of Web content on the mobile device requiring expert interpretation and hence not automated. Apart from this, there are some other guidelines which were found in expert evaluation, though not covered by any tool, such as “3.3 scrolling,” “3.5 graphics,” and “3.6 color.”

Apart from these guidelines mentioned in Overall Behavior, some were neither covered by tools nor by experts; for example, “3.1 page content” which states “ensure content is suitable for use in a mobile context, use clear and simple language, limit content to what the user has requested” is subjective and requires human judgement or “2.4 navigation mechanism” require real interaction.

Thus, the scope of this study is those guidelines whose violations are found by the tools and/or expert evaluation and are automatable. The guidelines whose at least one true violation is reported by all the tools include “2.8 refreshing, redirection and spawned window,” “4.2 frames,” “4.5 Non text items,” “4.6 Image size” “4.8 Measures,” “4.12 character encoding.”

A total of 8200 violations were found on all the Web sites by expert evaluation. However, if checkpoint “2.5 Access Keys” violation, this figure comes down to 3945. The most frequently violated mobile accessibility guidelines were “2.5 Access keys,” “2.8 Refreshing, Redirection and Spawned Windows,” “3.2 page size,” “3.6 Color,” “4.5 Non text items,” “4.6 Image size,” “4.7 Valid markup,” “4.12 Content types,” and “5.3 Labels for form controls” with 4255, 134, 38, 313, 962, 263, and 63 errors, respectively. The number of errors according to Navigation and Links, Page Layout and Content, Page Definition and User Input categories was 5389, 841, 1893, and 77, respectively.

### 7.1 Coverage

Coverage is one of the measures for evaluating the effectiveness of the tool. It is the extent to which a tool detects the true positive types with respect to true positives types identified by expert evaluation. Coverage is expressed as a percentage. True positives are the actual violations identified. A guideline is said to be covered by a tool if at least one true violation or true positive is found by the tool. Out of a total of 39 mobile Web accessibility guidelines, a maximum of 14 were found to be covered by the tools and 22 by expert evaluations. The coverage of an accessibility guideline is found by manually checking and confirming it by comparing the violation reported by the tool to violation detected by the expert evaluations. It was found that only 45–68 % were covered by automated tools. *Taw* and *mobileOK* report at least one true violation in 63 % of the guidelines while *EvalAccess* mobile reports violations in

**Table 1** Total number of guidelines covered by tools according to different a category

Cat.	Taw	EA	MOC	mobiReady	Total
NnL	3 60 %	2 40 %	3 60 %	3 60 %	5
PLnC	1 25 %	1 25 %	1 25 %	2 50 %	4
PD	10 90 %	5 45 %	10 90 %	9 81.8 %	11
UI	0	2	0	1	2
Total	14	10	14	15	22

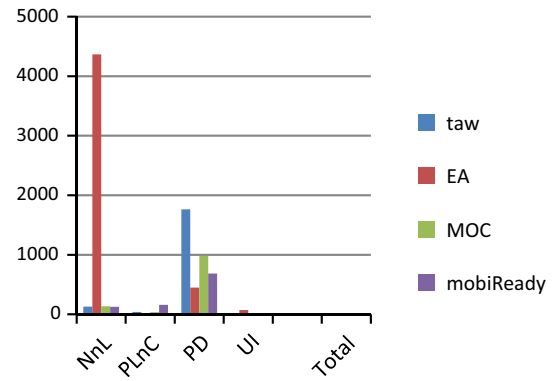
MOC mobileOK checker, EA EvalAccess mobile, NnL navigation and links, PLnC page layout and content, PD page definition, UI user interface

45 % of mobile accessibility guidelines. mobiReady reported coverage of 68 %. The number of guideline violations found, according to four groups in which mobile Web accessibility guidelines are categorized, Navigation and links, Page Layout and Content, Page definition, and User interface, was 5, 4, 11, 2, respectively. Coverage across these groups ranged from 40 to 60 % for Navigation and links, to 25 % for Page Layout and Content, 45–90 % for page definition, and 0–100 % for User Interface. Page Definition mobile Web accessibility guidelines were largely covered by taw and mobileOK tools, while the User Interface category was only covered by the EvalAccess Mobile tool. The other two categories were covered almost equally by all the four tools (Table 1).

### 7.2 Completeness

Completeness measures the number of instances of true mobile Web adequacy guideline violations found by tools with respect to the actual number of violations found in the Web sites by the experts. The total number of true violations as found by the experts in the study were 8200 including “2.5 Access keys” checkpoint violation and 3945 excluding “2.5 Access keys” checkpoint violations. The completeness ranges from 14 % by MobileOK to 59 % by EvalAccess. The number of true positives across four mobile Web accessibility guidelines category and tools is shown in Fig. 1.

It can be seen that most of the tools performed well for the Page definition. However, the EvalAccess tool performed well in Navigation and Links too. The reason for the high score of EvalAccess tool for Navigation and Links is that it has reported access key guideline violations which no other tool reported. One issue in this case to consider is how many access keys should be assigned and to which type of links, i.e., whether all the links that are not assigned access keys should be considered as a guideline violation.



**Fig. 1** Completeness per tool and category

**Table 2** Total number of true positives found by each of the tools

	TP	Taw	EA	MOC	mobiReady
Total	8200	1931 23 %	4899 59 %	1149 14 %	979 11.2 %
NnL	5389	130	4366	134	127
PLnC	841	38	14	35	158
PD	1893	1763	448	980	686
UI	77	0	71	0	7
Fn	–	6269	3301	7051	7222
Total-AK	3945	1931 48.9 %	644 16.3 %	1149 29.1 %	979 24.8 %
Fn(-AK)	–	2014	3301	2796	2967

Ak access keys, Fn false negatives, NnL navigation and link, PLnC page layout and content, PD page design, UI user interface, AK access key, TP true positives, EA EvalAccess mobile, MOC mobileOK checker

If the “2.5 Access keys” guideline is not taken into account for violations, then the total number of true positives of EvalAccess mobile tool becomes 644 and its completeness drops to 16 %, while completeness of taw, MobileOK checker and mobiReady increases to 48, 29, and 24 %, respectively (Table 2).

To find out which result was closer to the actual true positives, the correlation between the actual true positives and the actual result given by the tools was computed. The correlation between the results of the tools was also computed (Table 3).

The result shows that taw and mobileOK checkers show significant correlation (.961) among them and almost equal correlation with the actual true positives found by the experts.

### 7.3 Correctness

Correctness is the measure of how effectively the tool minimizes the number of errors reported by mistake, i.e.,

**Table 3** Correlation between the output of different tools and actual true positive by the experts

		Taw	EA	MOC	MobiReady	ATP
Taw	Correlation Coefficient	1.000	.394*	.961**	.696**	.549**
	Sig. (2-tailed)		.023	.000	.000	.001
EA	Correlation Coefficient	.394*	1.000	.455**	.616**	.391*
	Sig. (2-tailed)	.023		.008	.000	.025
MOC	Correlation Coefficient	.961**	.455**	1.000	.744**	.586**
	Sig. (2-tailed)	.000	.008		.000	.000
MobiReady	Correlation Coefficient	.696**	.616**	.744**	1.000	.557**
	Sig. (2-tailed)	.000	.000	.000		.001
ATP	Correlation Coefficient	.549**	.391*	.586**	.557**	1.000
	Sig. (2-tailed)	.001	.025	.000	.001	

ATP actual true positives, EA EvalAccess mobile, MOC MobileOK checker

\* Correlation is significant at the .05 level (2-tailed)

\*\* Correlation is significant at the .01 level (2-tailed)

**Table 4** Number of false positives and ratio of false positive and actual number of violations (not including guidelines “2.5 Access keys”)

	FP	Incorrectness (%)
Taw	1868	49
EA	813	55
w3c	1299	53
mobiReady	1400	58

EA EvaAccess

how well the tools reduce the number of false positives. False positives are errors reported by tools, which are not true errors. The instance of guideline violation reported by the tool is said to be a false positive if on comparing it with actual violation as detected by the experts is found that it is not a violation. The correctness is computed by subtracting incorrectness from 100. Incorrectness is computed as a ratio of false positives to the sum of false positives and true positives, i.e.  $(fp/(fp + tp))$ . It was found that the tools reported correctness ranging from 42 to 51 % (Table 4).

#### 7.4 Statistical data analysis

To further investigate the effectiveness of the tools, the data collected were classified into true positives, false negatives and false positives. The true positives, false negatives, and false positives reported by these tools were analyzed statistically to investigate whether there was any difference in the behavior of these tools and draw appropriate conclusions. The SPSS tool was used for the statistical analysis.

In order to investigate whether the four tools behave the same or different when detecting true positives, false negatives, and false positives, the following hypotheses were formed for statistical analysis:

**H01** All tools behave the same when detecting true positives. All four tools have equal medians for true positives;

**H02** All tools behave the same for false negatives. All four tools have equal medians for false negatives;

**H03** All tools behave the same for false positives. All four tools have equal medians for false positives;

To investigate the type of test to be conducted on the data, it was checked for normality. The normality of the data can be checked by a Kolmogorov–Smirnov and the Shapiro–Wilk test. The Kolmogorov–Smirnov test is used for large samples as large as 2000, and the Shapiro–Wilk test is used for small samples. Since the size of the data was small (<50 samples), the Shapiro test was used to test for normality (Table 5).

The result shows Sig. < .05 for Shapiro–Wilk test; therefore, the data are not normally distributed.

There are various nonparametric tests of significance for non-normal data. The Kruskal–Wallis test is a nonparametric test used to compare three or more samples. It is used to test the null hypothesis that all populations have identical distribution.

Here since there are four tools, the Kruskal–Wallis test is used to investigate whether all tools behave the same or differently when detecting true positives, false positives, and false negatives (Table 6).

The result of the Kruskal–Wallis test shows that there is a statistically significant difference between the true positives of tools (sig. < .05). So, the null hypothesis that all tools behave the same when detecting true positives; i.e., all four tools have equal medians for true positives is rejected. However, for false negatives and false positives significance is greater than .05 (Sig. > .05). Hence, the null hypothesis in both cases cannot be rejected.



**Table 5** Normality test results

	Shapiro–Wilk		
	Statistic	df	Sig.
Tests of normality			
TP	.775	44	.000
FN	.680	44	.000
FP	.887	44	.000

*TP* true positives, *FN* false negatives, *FP* false positives

**Table 6** Kruskal–Wallis test statistics

	TP	FN	FP
Chi-square	9.529	1.755	1.986
df	3	3	3
Asymp. sig.	.023	.625	.575

In order to further investigate whether all four tools behave differently or some tools behave in the same manner for true positives, a Mann–Whitney *U*-test was conducted. Mann–Whitney *U*-test is a nonparametric test performed on a non-normal distribution and is used to compare differences between two independent groups.

The following hypotheses were formed for pairwise comparison of tools for true positives:

- H#1** There is no difference between the true positives of law and EvalAccess mobile tools;
- H#2** There is no difference between the true positives of law and mobileOK checker tools;
- H#3** There is no difference between the true positives of law and mobiReady tools;
- H#4** There is no difference between the true positives of EvalAccess mobile and mobileOK checker tools;
- H#5** There is no difference between the true positives of EvalAccess mobile and mobiReady tools;
- H#6** There is no difference between the true positives of mobileOK checker and mobiReady tools (Table 7).

**Table 7** Mann–Whitney *U*-test statistics at  $\alpha = .05$

Pair of tools	Sig.	Mann–Whitney <i>U</i> -test
Taw and EvalAccess mobile	.005	18
Taw and mobileOK checker	.237	42.5
Taw and mobiReady	.071	33
EvalAccess mobile and mobileOK checker	.057	31.5
EvalAccess mobile and mobiReady	.131	37.5
mobileOK checker and mobiReady	.533	51

The result of the Mann–Whitney *U*-test shows that the significance value is  $<\alpha = .05$  only in case of law and EvalAccess mobile, thus showing that there is a difference in behavior of law and EvalAccess mobile in detecting true positives. Hence, the null hypothesis ( $H\#1$ ) is rejected. However, the rest of the pairs—law and mobileOK checker, law and mobiReady, EvalAccess mobile and mobileOK checker, EvalAccess mobile and mobiReady, and mobileOK checker and mobiReady do not show any significant difference in detecting the true positives as  $\text{sig.} > .05$  in all cases. Hence, the hypotheses  $H\#2$ ,  $H\#3$ ,  $H\#4$ ,  $H\#5$ , and  $H\#6$  are rejected.

### 8 Issues with MWBP and their interpretation with the tools

This section investigates and discusses the answer to the research question 3.

The W3C.org Web site states: “Being mobileOK is neither a guarantee that the Web document maybe rendered correctly by all mobile devices, nor insurance that the user experience was correctly addressed.” This statement says that even if the site validates MWBP, this does not mean it looks good or functions correctly across all the mobile devices.

The MWBP guidelines were formed in 2008 and extended in 2009. Due to the growth of mobile technology in terms of platforms, screen sizes, and high bandwidth available, there is a need to update and modify some MWBP guidelines. Some of these guidelines are discussed in this section.

One of the violations reported by the tools was “4.12 Character Encoding.” A character encoding tells the computer how to interpret the character. The mobile Web accessibility guideline “4.12 Character Encoding” says that encoding UTF-8 for attribute charset should be specified for the Web page. This is the issue of machine interpretation of the guideline. The HTML5 specification requires that whole meta element fit in the first 1024 bytes, of the document, and must be included at the top of the head element as `<meta Charset=“UTF-8”>`. The charset attribute is new in HTML5 and replaces the need for a

pragma directive of HTML4 specified as `<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">`. On studying the source code for which the tools reported this error, it was found that the tools reported character encoding guideline violation if the Web sites had charset encoding specified in the pragma directive of HTML4 not in HTML5.

TabIndex is an attribute in HTML whose value determines and customizes the tab order or navigation order of elements in HTML. While evaluating the Web sites, guideline “5.2 Tab Order” has not been considered as a violation in the present study if the tabIndex attribute is not present. However, it is checked that if it has been used in the Web page, it follows a logical order and used consistently. The guidelines “5.2 Tab Order” states “Create a logical order through links, form controls and objects”; i.e., if the tab key is used to move the cursor, the cursor movement should follow the structural order of the page elements so that user can select and activate them. However, it does not say that absence of tabIndex key violates the guidelines. In almost all circumstances, Web AIM (Web Accessibility in Mind) recommends against using tab Index because it can unintentionally create an illogical tab order [31]. Web AIM is a non-profit organization which provides knowledge, technical skills, tools, organizational leadership strategies, and vision so that organizations can make their Web content more accessible to people with disabilities [32].

Accesskey specifies a shortcut to activate or focus an HTML element. The W3C recommends “2.5 Access keys” the implementation of the accesskey attribute to enable users to select the appropriate key on their keyboards and navigate to a particular link. Accesskeys can also be useful to people who have no problems controlling the mouse and clicking on links. This study found that none of the Web sites under investigation had implemented access keys. This key has been largely underused due to two major flaws. The first problem is that visitors to a Web site have no way of knowing that accesskey attributes has been assigned to linked elements. Second, which accesskeys are assigned to which elements is also not known. Another problem is that many standards-compliant browsers still do not support accesskeys at all. Access keys are also irrelevant for touchscreen devices and, most importantly, adding them to pages can risk making the site less accessible by clashing with browser/screen reader defaults [33].

In the present study, it was found that the average page size of all the Web pages under study was 61.94 kb. This is the size of HTML source code excluding style sheet, external JavaScript, and embedded objects. The total average page size in this study, including source code, text, images, is 909.75 kb. According to MWBP guideline “5.3.2 Page size,” the page size should not exceed 20 kb. Today, the average page size served on mobile is 897 kb

[34]. According to an AKMAI report, the average mobile connection speeds (aggregated at a country level) in the first quarter of 2015 ranged from a high of 20.4 Mbps in the United Kingdom down to a low of 1.3 Mbps in Vietnam [35]. According to the above, a page of 897 kb served on a mobile device would take from .04 to .67 s to load. Therefore, the high resolution of smartphone screens and fast Internet speed requires modification in the recommendation of 20 kb page size.

Another guideline “5.4.11 Content Types” states “Send content in a format that is known to be supported by the device.” It implies that if a document contains images that are not in GIF or JPEG format, these are not correctly served by the Web server; i.e., images in formats other than GIF or JPEG may not work on some mobile browsers, and therefore, using PNG images is a violation of MWBP guideline. Today, PNG images are well supported by all mobile devices.

Another important issue is that the MWBP do not specify the degree of importance of various guidelines or severity of their violations as mentioned in WCAG guidelines. So, the tools except mobileOK do not distinguish between the guidelines that must be implemented from those that may be implemented.

Also, not all the MWBP guidelines are automated. For example, the ‘Overall Behavior’ category guidelines include some general principles regarding delivery to mobile devices.

## 9 Intra- and inter-reliability

It is observed by Diaper et al. [36] that different tools give different results even while assessing the same set of guidelines. This section is focusing on intrareliability and inter-reliability of the tools.

Intrareliability of automated tools is the ability of the tools to give consistently the same results for the same set of tests and data under the same conditions. It is used to determine the stability of the automated tools. Six pages selected randomly from the pages under study were tested twice with each tool to establish the stability of the tools. The result demonstrated that the tools were highly stable.

Inter-reliability refers to the extent to which two or more independent coders agree on the coding of the content of interest in an application of the same coding scheme. In this section, the agreement between tools with respect to MWBP guidelines is tested using Krippendorff’s alpha ( $\alpha$ ) Reliability Coefficient, a reliability measure proposed by Klaus Krippendorff. It is a reliability coefficient developed to measure the agreement among observers, coders, judges, raters, or measuring instruments, draw distinctions among typically unstructured phenomena, or assign computable values to them [37]. It can be used regardless of the

number of observers, levels of measurement, sample sizes, and presence or absence of missing data [38]. The general form of alpha computation is as follows:

$$\alpha = 1 - \frac{D_o}{D_e} \quad (1)$$

where  $D_o$  is the observed disagreement among values assigned to units of analysis and  $D_e$  is the disagreement one would expect when the coding of units is attributable to chance rather than to the properties of these units.  $\alpha = 1$  indicates perfect reliability;  $\alpha = 0$  indicates the absence of reliability. Units and the values assigned to them are statistically unrelated,  $\alpha < 0$  when disagreements are systematic and exceed what can be expected by chance.  $\alpha$  evaluates reliability one variable at a time.

The macro for computing Krippendorff's alpha is given by Hayes which can be downloaded and installed in SPSS [39]. In this study, the macro for computing Krippendorff's alpha was downloaded and installed in SPSS. Krippendorff's alpha measures agreements for nominal, ordinal, interval, and ratio data.

Checkpoints chosen for the analysis of agreement between the tools are those that are covered by at least two tools under study.

The results of the computations are tabulated in Table 8 which shows the result of the level of agreement between all four tools. It also tabulates the level of agreement by taking three tools at a time, omitting one of the tools to see whether the level of agreement increases or decreases.

The results show substantial discrepancies were found in the outcome of the tools while evaluating the same set of guidelines. There was perfect agreement between tools for guidelines "4.10 Minimize." This may be because only two tools *taw* and *mobileOK* detected it. The guideline was found to be violated by only one page among all the pages. The complete absence of agreement was found for guidelines "4.15 cache header" violation detection. The tools showed a high level of agreement for the guidelines "2.8 Refreshing, Redirection and Spawned Windows" (.8041), "4.2 frames" (.742), and "4.4 Tables" (.7416). The table also shows the result of an agreement between the tools if one out of four tools is excluded from the computations.

To find out the level of agreement between the actual true positives found by the experts and each tool, pairwise Krippendorff's alpha was computed. The result shows that *taw* and *mobileOK* checker show the similar value of alpha with actual true positives. This reflection was also observed in the correlation coefficient of the results found by *taw* and *mobileOK* checker.

The discrepancies may have arisen due to the subjective nature of the guidelines or incorrect interpretation of guidelines by the tools (Table 9).

## 10 Threat to validity and limitations

The threat to the validity of this research lies in small data samples, limitations of interpretation of the guidelines with the tools and human experts, the wide variety of hardware and software for mobile devices. Since the Web sites were tested online, difference in the time of testing of the same Web site by different experts may pose some threat to the validity of the study. Owing to responsive design, many Web sites respond differently to different devices. As, for example, the following table (Table 10) shows some key differences between the different versions of the Facebook "Home" page served to different devices after logging into facebook.com.

Though the automated tools provide a quick and easy way of evaluating the mobile accessibility of Web sites, another threat to the cogency of the study is that the MWBP guidelines were last updated in 2009, and there is a need for revising them in the context of mobile devices available today. All the tools of Web adequacy available are based on MWBP.

WCAG 2.0, an updated version of WCAG 1.0 for improving adequacy of Web content, also supports some guidelines for mobile Web content too. There is some overlap between WCAG 2.0 and MWBP guidelines, though all MWBP guidelines cannot be directly mapped to WCAG 2.0. The relationship between MWBP and WCAG 2.0 can be found in [40]. There is also a lack of accessibility evaluation tools that support both WCAG 2.0 and MWBP. Developers utilize evaluation tools based on MWBP to enhance mobile Web adequacy to provide a reasonable experience in the mobile Web. The relevance and appropriateness of mobile Web guidelines are also discussed by Clegg-Vinell et al. [41].

## 11 Conclusion

It was observed that *EvalAccess* performed well in terms of completeness while *taw* and *MobileOK* did well in coverage and correctness. Automatic tools do not cover all guidelines since not all the guidelines can be automated. It was observed that only 22 guidelines were covered by tools and/or expert evaluations. The study shows that coverage of the automated tools lies between 45 % (*EvalAccess*) and 68 % (*mobiReady*). The completeness ranged from 14 % by *MobileOK* to 59 % by *EvalAccess* mobile while correctness was found between 42 and 51 %.

Statistical analysis revealed that there was no difference in reporting the false positives and false negatives between the tools. However, in the case of true positives, statistical investigation showed that only *taw* and *EvalAccess* had a difference in reporting true positives while the reporting of

**Table 8** Krippendorff's alpha ( $K\alpha$ ) computations between the results given by taking all tools together and three tools at a time

CP	$K\alpha$ (all tools)	$K\alpha$ (1,2,3)	$K\alpha$ (1,3,4)	$K\alpha$ (2,3,4)	$K\alpha$ (1,2,4)
2.7	.0359	.0242	.3804	-.2384	-.2384
2.8	.8041	.9811	.7469	.7470	.7419
3.5	-.1681	-.4133	-.5284	-.1681	-.1371
4.2	.742	1	.6394	.6394	.6394
4.4	.7416	.9976	.7416	.6053	.6095
4.5	.5294	.5295	.7334	.3954	.4515
4.6	.6241	.6391	.6979	.5656	.6013
4.7	-.0944	-.3617	-.0944	.1442	-.2685
4.8	.1408	.1681	.2487	.0664	.0552
4.10	1	-	-	-	-
4.11	.3008	-	-	-	-
4.12	.2089	.1567	.5938	.0574	.0741
4.15	0	-	0	-	-
5.1	-.2348	-	-	-	-

CP Checkpoint, 1 Taw, 2 EvalAccess, 3 mobileOK, 4 ReadyMobi

**Table 9** Krippendorff alpha between actual true positives with each tool

CP	Taw	EA	MOC	mobiReady
2.7	.507	-.024	.507	.335
2.8	.814	.7987	.8207	.7267
2.9	.77		.4464	.335
3.5		.0328	-.179	.2992
4.2	1	1	1	-.244
4.4	.835		.8341	.4037
4.5	.866	.411	.7902	.5479
4.6	.806	.6936	.717	.7749
4.7	.37		-.211	-.103
4.8	.29	-.101	.1936	.4544
4.1	.129		.438	

CP Checkpoint, EA EvalAccess mobile, MOC MobileOK checker

the rest of the tools did not show any significant difference. It was observed from the study that all the tools under study were almost equally effective in evaluating mobile Web accessibility. As the results demonstrate, the tools were highly stable though substantial discrepancies were found in the outcome of the tools while evaluating the same lot of guidelines. The Krippendorff's alpha test revealed that Taw and mobileOK checker tools showed similar results while evaluating the MWBP guidelines in most cases.

Though statistical tests demonstrate that all the tools under study show similar behavior for coverage, completeness, and correctness, still Taw and MobileOK show the better coverage, completeness, and correctness as compared to others.

However, no tools are self-sufficient for adequacy evaluation. For a more precise evaluation, multiple tools along with human judgment need to be applied.

It was felt during the study that it may be more helpful if mobile Web adequacy evaluation tools in addition to reporting the details of the guidelines violated by the Web site may also incorporate the details about which guidelines are being adhered to by the Web site and to what extent. For example, if guideline X has been followed 4 out of 6 times, the tool must reflect this in the report. In this way, developers would be able to know whether the guideline has been followed by the Web site and reported by the tool, followed by the Web site, but missed by the tool or whether the guideline has not been followed by the Web site and not detected by the tool as a violation. This would save lots of effort of the developer working toward Web adequacy irrespective of the devices.

Also, different smart phones support different platforms, each having their own security and accessibility features. This open application model may pose some security threats for organizations. The voice control functions are usually available in English, using an American English accent, and therefore cannot accurately take the voice commands in a different accent. The screen readers are also mostly available in English, thus resulting in a limited user group being addressed. There is a need of investigation and study of major issues for enabling the Mobile Web in Indian languages and the extent of effectiveness of automatic mobile Web evaluation tools in the context of Indian Web sites particularly in the context of the various regional languages used, where pronunciation may also vary with region. The same spellings have a different meaning, and the same words have different spellings.

Mobile adequacy is no longer about creating a Web site compliant with MWBP guidelines. The mobile Web adequacy evaluation tools must be developed taking device

**Table 10** Difference in the home page rendering of [www.facebook.com](http://www.facebook.com) by different mobile devices. (Source <https://mobiforge.com/research-analysis/anatomy-a-mobile-web-experience-facebook-com>)

	iPhone	BBC	Nokia 6300	Simple site
HTML size (Kb)	31	13	11	3
Images size (Kb)	17	33	33	–
Java script size (Kb)	62	–	–	–
CSS size (Kb)	15	Inline	8	8
Total page weight (Kb)	124	46	53	11
No. of HTTP requests	14	13	16	5
Doctype served	XHTML Mobile 1.0	XHTML Mobile 1.0	XHTML Mobile 1.0	HTML Mobile 1.0
CSS usage	extensive	Minimal	minimal	Minimal
Data URIs	✓	✗	✗	✗

BBC black berry curve

and platform characteristics into considerations. Due to the rapid pace of mobile technology development in mobile device features and responsive design improvements, many of the guidelines need to be revised.

## Appendix

See Table 11.

**Table 11** List of webpages considered in the study

Sno	Web pages
1	<a href="http://www.businessweek.com/">http://www.businessweek.com/</a>
2	<a href="http://www.businessweek.com/global-economics">http://www.businessweek.com/global-economics</a>
3	<a href="http://www.caranddriver.com/">http://www.caranddriver.com/</a>
4	<a href="http://buyersguide.caranddriver.com/acura">http://buyersguide.caranddriver.com/acura</a>
5	<a href="http://www.usa.gov">http://www.usa.gov</a>
6	<a href="http://www.fueleconomy.gov/feg/gasprices/states/index.shtml">http://www.fueleconomy.gov/feg/gasprices/states/index.shtml</a>
7	<a href="http://www.dallasnews.com/">http://www.dallasnews.com/</a>
8	<a href="http://www.dallasnews.com/news/">http://www.dallasnews.com/news/</a>
9	<a href="http://mobile.uk.msn.com/device/default.aspx">http://mobile.uk.msn.com/device/default.aspx</a>
10	<a href="http://mobile.uk.msn.com/device/ent/?ocid=naventertainment">http://mobile.uk.msn.com/device/ent/?ocid=naventertainment</a>
11	<a href="http://mobile.technorati.com/">http://mobile.technorati.com/</a>
12	<a href="http://mobile.technorati.com/women/">http://mobile.technorati.com/women/</a>
13	<a href="http://m.bestbuy.com/m/e/digital/">http://m.bestbuy.com/m/e/digital/</a>
14	<a href="http://m.bestbuy.com/m/e/digital/popular">http://m.bestbuy.com/m/e/digital/popular</a>
15	<a href="http://m.espn.go.com/wireless/index?w=1dnl&amp;i=MOBI">http://m.espn.go.com/wireless/index?w=1dnl&amp;i=MOBI</a>
16	<a href="http://m.espn.go.com/nfl/">http://m.espn.go.com/nfl/</a>
17	<a href="http://www.cnet.com/">http://www.cnet.com/</a>
18	<a href="http://news.cnet.com/">http://news.cnet.com/</a>
19	<a href="http://www.lonelyplanet.com/">http://www.lonelyplanet.com/</a>
20	<a href="http://shop.lonelyplanet.com/">http://shop.lonelyplanet.com/</a>
21	<a href="http://www.accuweather.com/en/au/australia-weather">http://www.accuweather.com/en/au/australia-weather</a>
22	<a href="http://www.accuweather.com/en/climate-change">http://www.accuweather.com/en/climate-change</a>

## References

- Kemp, S.: Digital, social and mobile worldwide in 2015. <http://wearesocial.net/blog/2015/01/digital-social-mobile-worldwide-2015/> (2015)
- Mobile and Desktop Collide: What Are You Doing About It? <http://www.mequoda.com/articles/digital-magazine-publishing/web-usage-prediction-when-mobile-and-desktop-collide/> (2015)
- Infographic: Mobile growth statistics. <http://www.digitalbuzzblog.com/infographic-2013-mobile-growth-statistics/> (2013)
- Importance of mobile web in higher education recruitment: A student's perspective. <http://www.bluetrainmobile.com/blog/the-importance-of-mobile-web-in-higher-education-recruitment-a-students-perspective/> (2013)
- The importance of mobile optimization. <http://www.getelastic.com/the-importance-of-mobile-optimization-infographic> (2013)
- Top 6 reasons your shopping cart abandonment rate is high on mobile. <http://www.mobify.com/insights/high-shopping-cart-abandonment-rate-on-mobile/> (2014)
- Ivory, M.Y.: Automated web site evaluation: researchers' and practitioners' perspectives. Springer Science and Business Media, 31 Oct 2003-Computers (2003)
- Vigo, M., Brown, J., Conway, V.: Benchmarking web accessibility evaluation tools: measuring the harm of sole reliance on automated tests. In: Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility W4A (2013)
- Mankoff, J., Fait, H., Tran, T.: Is your page accessible? A comparative study of methods for assessing web page accessibility for the blind. In: Proceedings of the SIGCHI conference on Human factors in computing system, CHI'05, pp 41–50 (2005)
- From WCAG 2.0 to MWBP: Making content that meets Web Content Accessibility Guidelines 2.0 also meet Mobile Web Best Practices. <http://www.w3.org/TR/mwbp-wcag/wcag20-mwbp.html> (2009)
- Yesilada, Y., Brajnik, G., Harper, S.: Barriers common to mobile and disabled web users. Inter. Comput. **23**(5), 525–542 (2011)
- Vigo, M., Aizpurua, A., Arrue, M., Abascal, J.: Evaluating the web accessibility for specific mobile devices. In: 5th International Cross-Disciplinary Conference on Web Accessibility, W4A 2008, pp 65–72 (2008)
- Bandeira, R., Lopes, R., Carriço, L.: Towards mobile Web accessibility evaluation. In: ETAPS, FOSS-AMA Workshop, Paphos, Cyprus (2010)
- Carriço, L., Lopes, L., Bandeira, R. Crosschecking the mobile web for people with visual impairments. In: Proceedings of the International Cross-Disciplinary Conference on Web



- Accessibility, W4A 2011, Hyderabad, Andhra Pradesh, India, March 28–29 (2011)
15. Carriço, L., Lopes, L., Bandeira, R., Isacker, K.V.: Towards mobile web accessibility: vision and challenges. In: Proceedings of the 1st International AEGIS Conference, Seville, Spain, October (2010)
  16. Vigo, M., Aizpurua, A., Arrue, M., Abascal, J.: Quantitative assessment of mobile web guidelines conformance. *Univers. Access Inf. Soc.* **10**(1), 33–49 (2011)
  17. Rabin, J., McCathieNevile, C.: Mobile Web Best Practices 1.0. W3C Mobile Web Initiative. <http://www.w3.org/TR/mobile-bp/> (2008)
  18. Brajnik, G.: Comparing accessibility evaluation tools: a method for tool effectiveness. *Univers. Access Inf. Soc.* **3**, 252–263 (2004)
  19. W3C: Web Content Accessibility Guidelines [online]. Recommendation, W3C. <http://www.w3.org/TR/WCAG10/> (1999)
  20. W3C: Overview of WCAG 2.0 Documents, [online] Web Accessibility Initiative. <http://www.w3.org/TR/WCAG20/> (2008)
  21. Mobile Web Best Practices. <http://www.w3.org/TR/mobile-bp/>
  22. Extended Guidelines for Mobile Web Best Practices 1.0. <http://www.w3.org/TR/2009/NOTE-mwbp-guidelines-20091020/>
  23. Owen, S., Rabin, J.: W3C mobileOK Basic Tests 1.0. W3C Mobile Web Initiative. <http://www.w3.org/TR/mobileOK-basic10-tests/> (2008)
  24. MobileOk pro test. Addendum to Mobile Web Best Practices. <http://www.w3.org/2005/MWI/BPWG/Group/TaskForces/mobileOKPro/drafts/ED-mobileOK-pro10-tests-20090923>
  25. Hackett, S., Parmanto, B.: Homepage not enough when evaluating web site accessibility. *Internet Res.* **19**(1), 78–87 (2009)
  26. Random Sampling. <https://explorable.com/simple-random-sampling>
  27. W3C mobileOk checker. <http://validator.w3.org/mobile/>
  28. EvalAccess. <http://supt07.si.edu/es/evalaccess2/mobile/index.html>
  29. Taw. <http://www.tawdis.net/>
  30. dotMobi, mobiReady. [http://ready.mobi/launch.jsp?locale=en\\_EN](http://ready.mobi/launch.jsp?locale=en_EN)
  31. Keyboard accessibility. <http://webaim.org/techniques/keyboard/tabindex>
  32. WebAIM. <http://webaim.org/>
  33. Accesskeys. <http://alistapart.com/article/accesskeys>
  34. Everts, T.: Want to Deliver the best possible user experience in 2015? Consider these 8 web performance resolutions. <http://blog.radware.com/applicationdelivery/applicationaccelerationoptimization/2015/01/8-web-performance-resolutions-for-2015/> (2015)
  35. Akamai report on Internet. <http://www.akamai.com/dl/soti/q1-2015-soti-fullreport-a4.pdf>
  36. Diaper, D., Worman, L.: Two falls out of three in the automated accessibility assessment of world wide web sites: A-Prompt v. Bobby. In: Johnson, P., Palanque, P., O'Neill, E. (eds.) *People and Computers XVII*, Proceeding of 17th Annual Human Computer Interaction Conference, pp. 349–363. Springer, Berlin (2004)
  37. Krippendorff, K.: Computing Krippendorff's Alpha-Reliability. Retrieved from [http://repository.upenn.edu/asc\\_papers/43](http://repository.upenn.edu/asc_papers/43) (2011)
  38. Hayes, A.F., Krippendorff, K.: Answering the call for a standard reliability measure for coding data. *Commun. Methods Meas.* **1**, 77–89 (2007)
  39. Hayes Macro for SPSS. <http://www.afhayes.com/spss-sas-and-mplus-macros-and-code.html>
  40. From MWBP to WCAG 2.0: Making content that meets Mobile Web Best Practices also meet Web Content Accessibility Guidelines 2.0. <http://www.w3.org/TR/mwbp-wcag/mwbp-wcag20.html>
  41. Clegg-Vinell, R., Bailey, C., Gkatzidou, V.: Investigating the appropriateness and relevance of mobile Web guidelines. In: Proceedings of the 11th Web for All Conference W4A'14 (2014)