Guideline-Based Evaluation of Web Readability

Abstract

Previous studies have explored the comparison of machine evaluation and manual evaluation methods in web page evaluation, in order to understand the readability and accessibility evaluation protocols and systems in the future. The experimental results show that the proposed models have their own advantages and disadvantages in different situations. However, due to the differences in size and interface design between mobile display and laptop display, the experimental results can not guarantee the same conclusion on mobile display. So I want to study the difference between reading the same page on a mobile device and reading the same page on a web page. I use semi-structured interview and questionnaire survey methods to obtain data, and modeling, and get the results similar to the network experiment. But because web page reading on mobile phone is more complex than that on Web page, it needs more human experts. These results lay a foundation for the design evaluation methods of mobile phones in the future.

ACM Reference Format:

. 2018. Guideline-Based Evaluation of Web Readability . In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/1122445.1122456

1 Introduction

This article explored the use of readability guidelines in manual and automatic webpage evaluation, to derive the insights for future readability and accessibility evaluation protocols and systems. The study collected and compared three pieces of data for a set of webpages: the ground truth readability scores, manual evaluation with readability guidelines, and automatic evaluation on guideline-related metrics.

In this paper, the linear mixed model is used as the main tool to input automatic variables into a series of linear mixed models, and the relationship between automatic evaluation and readability variables is discussed.

The traditional manual evaluation method based on guide has some problems, for example, some lengthy and simple checklists greatly increase the unnecessary manual work. The analysis shows that the use of automatic evaluation can greatly alleviate these problems without reducing the evaluation effect. However, completely using automatic evaluation instead of manual evaluation will greatly increase the error probability in some cases. We believe that the future web page evaluation will be based on the combination of manual evaluation and automatic evaluation.

This paper studies 39 web page readability criteria and their application in automatic evaluation. Automatic assessment seems to help address several problematic aspects of the web page, including low-level visual aspects where manual assessment often leads to disagreements. However, experts still need to manually apply guidelines that cannot be used successfully by automatic assessment, such as content-based understanding and interpretation.

Due to the differences in size and UI design between mobile displays and laptop displays, people may read the same page differently on mobile devices and computers. In order to understand the differences between mobile reading and online reading, this study conducted a semi-structured interview and questionnaire survey among different ages and special groups.

This article contributes to a new study of the difference between people reading the same page and a web page on a mobile device. In this paper, semi-structured interview and questionnaire survey methods are used to obtain data and model. The experimental results are consistent with the experimental results of web page. When evaluating web pages according to specific standards, the performance of the algorithm on human pages is better than that of human experts, especially those related to the low-level functions of the readability and text format. However, multiple standards still need human judgment and understanding and interpretation of network content; The difference is that mobile reading is faced with more complex and changeable situations, and the manual intervention will become more frequent and necessary. These results are helpful to classify the guidelines and lay the foundation for future design evaluation methods.

Woodstock '18, June 03–05, 2018, Woodstock, NY

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00

https://doi.org/10.1145/1122445.1122456

2 pre-study

Due to the differences in size and interface design between mobile phone monitor and laptop monitor, people may read the same page differently on mobile phone and computer. In order to further understand the differences between mobile reading and online reading, we conducted a questionnaire survey on different user groups. Because our goal is to find out whether the results of mobile reading are similar to those of online reading, we hope to obtain some user data through questionnaire survey.

We still use some simple and practical indicators to help us understand how users read mobile phone pages, such as the fixation time and fixation times of text and non text obtained by manual timing and counting. Based on this index, we need to collect and compare three parts of data on a group of mobile phone pages: the actual reading situation, manual evaluation with readability guide, and guidance of automatic evaluation related indicators. We can collect multiple reports in a short period of time, view and analyze the report content, and get the difference between mobile reading and web reading.

We can use the reports we get to analyze how people read the same webpage on different devices. The results may be similar, which means that the impact of different devices on people's reading is negligible. If the results are very different, it means that different devices still have a great influence on people's reading. Does this influence come from the difference in UI design? Or the difference in page size? Or a difference in interaction design? Further investigation is needed.

At the same time, since web users cover both people with dyslexia and ordinary users, we should include both groups for survey users. When designing questions, we should put the dyslexia in the first place. At the same time, the results should be counted separately and given With different weightings, the previous survey shows that about 0.07 of users have dyslexia (note 22 above). In addition, readability is of greater significance to this type of user itself. After the results of the two groups are separately counted, it is proposed to The ratio of 8:2 is used as a reference for the final item. At the same time, the two groups should be considered separately for issues that have huge differences.

2.1 Ergonomics

In order to better allow users to use the web, we must first have sufficient knowledge of the users themselves. Only by roughly dividing the user's reading ability and having a general understanding of the user's reading habits can it be better for users to read web pages. Therefore, ergonomic introduc- tion is essential. This is a subject that includes knowledge of anatomy, physiology, and psychology, and studies the in- teraction between humans and machines to make people work most efficiently. It has been defined as an independent subject for more than 40 years.

And its development lies in the military use of weapons of war during World War II. Today, in order to apply the results of ergonomics to mobile phone manufacturing. At the same time, combined with ordinary people's reading on the web, the basic factors we consider in- clude: page layout, reading fonts, and the difficulty of words; in view of the recent popularity of converting text to speech, noise should also be considered.

2.2 Cognitive psychology

Cognitive psychology mainly studies people's psychological activities by studying the logic of the process of input and output judgment. Its core is to study how people understand things, including attention, perception, memory and thinking. Human perception is not only about seeing, listening, smelling and mocking, but also being accepted by human body and brain through image, sound, smell and other stimuli, and compared with memory. In the research of web usability, it is not limited to the cognitive model of visual image. It is very beneficial for some dyslexics to broaden the way of cognitive formation by stimulating factors such as sound and smell.

2.3 Research significance

Data shows that among the six billion people in the world today, at least 4.1 billion people are using mobile phones. It can be said that mobile phones, as the most commonly used devices, have become the most important tool for work and entertainment. When the diverse applications of mobile phones cover people's lives and work, they themselves transform from a single communication device to a mobile processing terminal with diversified functions. In the process of this transformation, human-computer interaction has a very important position. Strengthening the convenience of mobile reading and expanding the potential mobile reading crowd is undoubtedly a direction of the development of human-computer interaction in the future, which will eventually promote designers to meet the needs of more users.

3 Related Work

In the article Evaluating the Usability of Homestay Websites in Malaysia Using Automated Tools, the author evaluated the usability of Malaysia homestay websites by using various automated tools, such as web analyzers (from website optimization) and Dead Link Checker tools. The author found some usability issues, such as (i) page size, (ii) broken links and (iii) download speed. But the author did not interview real users. It would be better if the author could understand their perception and experience of website interaction.

For people with aphasia, written medical resources must be cognitively accessible, accurate and easy to understand. In the article Evaluation of the readability, validity, and user-friendliness of written web-based patient education materials for aphasia, the author objectively evaluates the accessibility of current written education materials for patients with aphasia and checks the readability of the obtained materials. The relationship between effectiveness and user-friendliness. For each webpage, use the following readability formula to evaluate readability: Flesch-Kincaid ease of reading, Flesch-Kincaid grade (F-K), Gobbledygook's simple metric, and FORCAST. In addition, the following measures were used to evaluate the effectiveness and user-friendliness of each web page: site, publisher, audience, timeliness (SPAT) and material suitability assessment (SAM). The author concludes that there is a significant difference between the readability, validity and user-friendliness of written information and the expected and current levels. But the author only considered the computer webpage reading, not the mobile phone reading, and this is the direction we want to study.

In the article Comp4Text Checker: An Automatic and Visual Evaluation Tool to Check the Readability of Spanish Web Pages, the author introduces the Comp4Text online readability evaluation tool. The tool can calculate the readability level of web pages based on classic language metrics (from sentence to sentence) and detect unusual words and abbreviations. In addition, it provides suggestions for solving readability issues and displays everything in a very intuitive way. With this tool, web designers and writers can improve their websites and make them easier for everyone to read and understand. However, this Comp4Text online readability evaluation tool is too mechanized, it can only be used for some ordinary pages, and cannot detect some special pages. Because the subjective opinions of humans have not been considered, this evaluation tool is not very capable of dealing with complex pages.

4 Approach

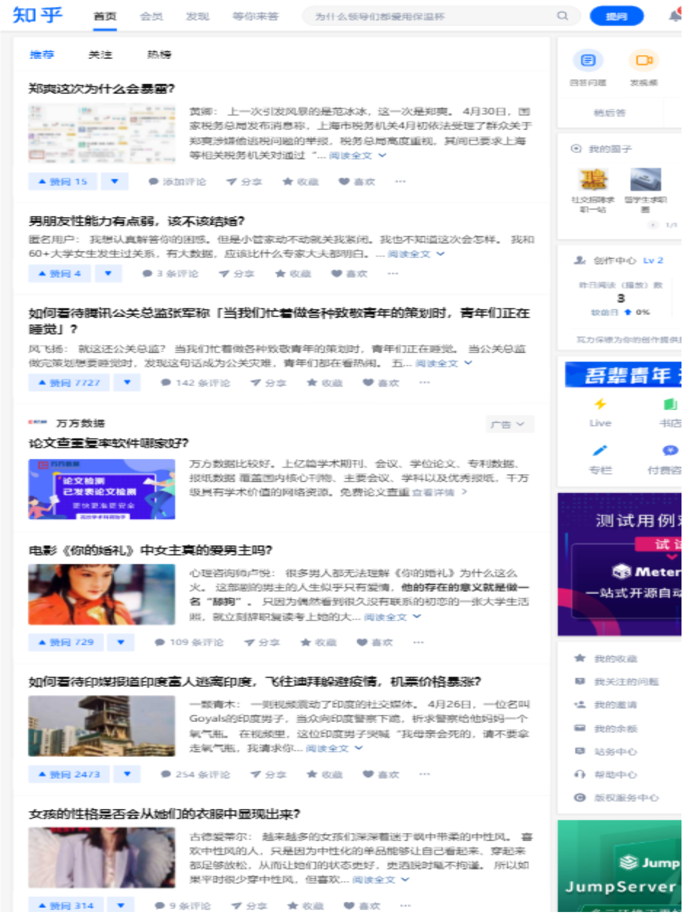
Since we need to judge whether there will be similar conclusions between mobile-based reading and computer-based reading, we did a control experiment in this experiment. The control group is based on the reading on the computer side, and the experimental group is based on the reading on the mobile phone.

The study collected and compared three pieces of data for a set of webpages: the ground truth readability scores, manual evaluation with readability guidelines, and automatic evaluation on guideline-related metrics.

4.1 Stimuli

We sampled 10 pages from news, entertainment and education websites. There is an article on health, research, new technology or education on the sample webpage-these topics are considered to be enough to attract children and adults to stay focused throughout the experiment and read through the text without skipping. We also ensured that these pages can be browsed on computers and mobile phones. Because we set up a control experiment, we need pages that can be browsed at the same time. For the mobile page, we have selected mobile phones of the same brand and model to ensure that the screen size and display method are consistent. For the web page, we also selected computers of the same brand and model to ensure that the screen size and display method are consistent.

The following two pictures are for the same page, the display difference between mobile browsing and computer browsing.



郑爽这次为什么会暴雷?

黄卿：上一次引发风暴的是范冰冰，这一次是郑爽。4月30日，国家税务总局发布消息称，上海市税务机关4月初依法受理了群众关于郑爽涉嫌偷逃税问题的举报，税务总局高度重视，其间已要求上海等相关税务机关对通过"... 阅读全文

▲ 赞同 15 ▼ 添加评论 分享 收藏 喜欢 ···

男朋友性能力有点弱，该不该结婚?

匿名用户：我想认真解答你的困惑。但是小管家动不动就关我案闭。我也不知道这次会怎样。我和60+大学女生发生过关系，有大数据，应该比什么专家大夫都明白。... 阅读全文

▲ 赞同 4 ▼ 3 条评论 分享 收藏 喜欢 ···

如何看待腾讯公关总监张军称「当我们忙着做各种致敬青年的策划时，青年们正在睡觉」?

风飞扬：就这还公关总监？当我们忙着做各种致敬青年的策划时，青年们正在睡觉。当公关总监做完策划想要睡觉时，发现这句话成为公关灾难，青年们都在看热闹。五... 阅读全文

▲ 赞同 7727 ▼ 142 条评论 分享 收藏 喜欢 ···

万方数据

论文查重复率软件哪家好?

万方数据比较好。上亿篇学术期刊、会议、学位论文、专利数据、报纸数据 覆盖国内核心刊物、主要会议、学科以及优秀报纸，千万级具有学术价值的网络资源。免费论文查重 查看详情 >

电影《你的婚礼》中女主真的爱男主吗?

心理咨询师卢悦：很多男人都无法理解《你的婚礼》为什么这么火。这部剧的男主的人生似乎只有爱情，**他的存在的意义就是做一名"舔狗"。**只因为偶然看到很久没有联系的初恋的一张大学生活照，就立刻辞职复读考上她的大... 阅读全文

▲ 赞同 729 ▼ 109 条评论 分享 收藏 喜欢 ···

如何看待印媒报道印度富人逃离印度，飞往迪拜躲避疫情，机票价格暴涨?

一颗青木：一则视频震动了印度的社交媒体。4月26日，一位名叫Goyals的印度男子，当众向印度警察下跪，祈求警察给他妈妈一个氧气瓶。在视频里，这位印度男子哭喊"我母亲会死的，请不要拿走氧气瓶，我请求你... 阅读全文

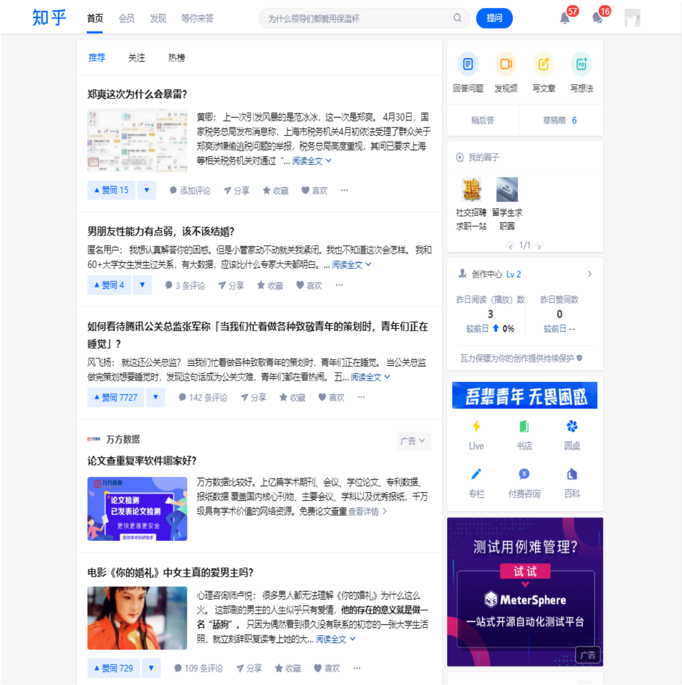
▲ 赞同 2473 ▼ 254 条评论 分享 收藏 喜欢 ···

女孩的性格是否会从她们的衣服中显现出来?

古德爱蒂尔：越来越多的女孩们深深着迷于飒中带柔的中性风。喜欢中性风的人，只是因为中性化的单品能够让自己看起来、穿起来都足够放松，从而让她们的状态更好，更洒脱时髦不拘谨。所以如果平时很少穿中性风，但喜欢... 阅读全文

▲ 赞同 314 ▼ 9 条评论 分享 收藏 喜欢 ···

Figure 1. Pages browsed on the mobile phone



郑爽这次为什么会暴雷?

黄卿：上一次引发风暴的是范冰冰，这一次是郑爽。4月30日，国家税务总局发布消息称，上海市税务机关4月初依法受理了群众关于郑爽涉嫌偷逃税问题的举报，税务总局高度重视，其间已要求上海等相关税务机关对通过 "... 阅读全文

▲ 赞同 15 ▼ 添加评论 分享 收藏 喜欢 ···

男朋友性能力有点弱，该不该结婚?

匿名用户：我想认真解答你的困惑。但是小管家动不动就关我紧闭。我也不知道这次会怎样。我和60+大学女生发生过关系，有大数据，应该比什么专家大夫都明白。... 阅读全文

▲ 赞同 4 ▼ 3 条评论 分享 收藏 喜欢 ···

如何看待腾讯公关总监张军称「当我们忙着做各种致敬青年的策划时，青年们正在睡觉」?

风飞扬：就这还公关总监? 当我们忙着做各种致敬青年的策划时，青年们正在睡觉。当公关总监做完策划想要睡觉时，发现这句话成为公关灾难，青年们都在看热闹。五... 阅读全文

▲ 赞同 7727 ▼ 142 条评论 分享 收藏 喜欢 ···

万方数据 广告

论文查重复率软件哪家好?

论文检测 已发表论文检测 更快更准更安全

万方数据比较好。上亿篇学术期刊、会议、学位论文、专利数据、报纸数据 覆盖国内核心刊物、主要会议、学科以及优秀报纸，千万级具有学术价值的网络资源。免费论文查重 查看详情

电影《你的婚礼》中女主真的爱男主吗?

心理咨询师卢悦：很多男人都无法理解《你的婚礼》为什么这么火。这部剧的男主的人生似乎只有爱情，**他的存在的意义就是做一名"舔狗"**。只因为偶然看到很久没有联系的初恋的一张大学生活照，就立刻辞职复读考上她的大... 阅读全文

▲ 赞同 729 ▼ 109 条评论 分享 收藏 喜欢 ···

Figure 2. Pages browsed on the computer

4.2 Readability Guidelines

We used a readability guide from a recent study that was collected with design and dyslexia experts to eliminate ambiguity, classify and review them [2]. Of the 47 guidelines listed that apply to individual pages, we have omitted ten guidelines, which seem to be violated by only a few pages. We also divided the two guides into four parts because they consist of two parts, so there are 39 guides in total.

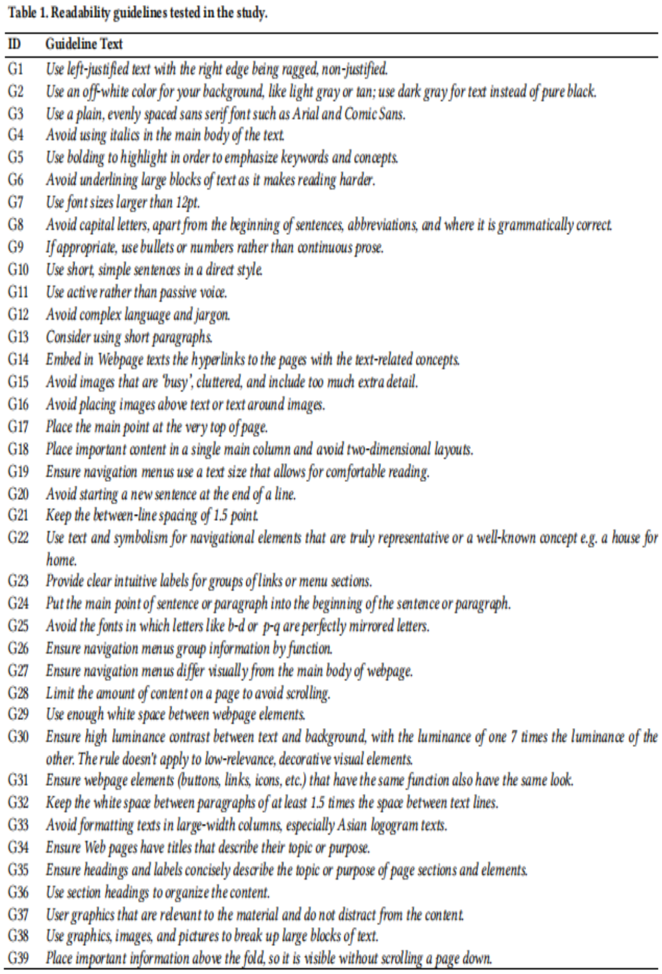


Table 1. Readability guidelines tested in the study.

ID	Guideline Text
G1	*Use left-justified text with the right edge being ragged, non-justified.*
G2	*Use an off-white color for your background, like light gray or tan; use dark gray for text instead of pure black.*
G3	*Use a plain, evenly spaced sans serif font such as Arial and Comic Sans.*
G4	*Avoid using italics in the main body of the text.*
G5	*Use bolding to highlight in order to emphasize keywords and concepts.*
G6	*Avoid underlining large blocks of text as it makes reading harder.*
G7	*Use font sizes larger than 12pt.*
G8	*Avoid capital letters, apart from the beginning of sentences, abbreviations, and where it is grammatically correct.*
G9	*If appropriate, use bullets or numbers rather than continuous prose.*
G10	*Use short, simple sentences in a direct style.*
G11	*Use active rather than passive voice.*
G12	*Avoid complex language and jargon.*
G13	*Consider using short paragraphs.*
G14	*Embed in Webpage texts the hyperlinks to the pages with the text-related concepts.*
G15	*Avoid images that are 'busy', cluttered, and include too much extra detail.*
G16	*Avoid placing images above text or text around images.*
G17	*Place the main point at the very top of page.*
G18	*Place important content in a single main column and avoid two-dimensional layouts.*
G19	*Ensure navigation menus use a text size that allows for comfortable reading.*
G20	*Avoid starting a new sentence at the end of a line.*
G21	*Keep the between-line spacing of 1.5 point.*
G22	*Use text and symbolism for navigational elements that are truly representative or a well-known concept e.g. a house for home.*
G23	*Provide clear intuitive labels for groups of links or menu sections.*
G24	*Put the main point of sentence or paragraph into the beginning of the sentence or paragraph.*
G25	*Avoid the fonts in which letters like b-d or p-q are perfectly mirrored letters.*
G26	*Ensure navigation menus group information by function.*
G27	*Ensure navigation menus differ visually from the main body of webpage.*
G28	*Limit the amount of content on a page to avoid scrolling.*
G29	*Use enough white space between webpage elements.*
G30	*Ensure high luminance contrast between text and background, with the luminance of one 7 times the luminance of the other. The rule doesn't apply to low-relevance, decorative visual elements.*
G31	*Ensure webpage elements (buttons, links, icons, etc.) that have the same function also have the same look.*
G32	*Keep the white space between paragraphs of at least 1.5 times the space between text lines.*
G33	*Avoid formatting texts in large-width columns, especially Asian logogram texts.*
G34	*Ensure Web pages have titles that describe their topic or purpose.*
G35	*Ensure headings and labels concisely describe the topic or purpose of page sections and elements.*
G36	*Use section headings to organize the content.*
G37	*User graphics that are relevant to the material and do not distract from the content.*
G38	*Use graphics, images, and pictures to break up large blocks of text.*
G39	*Place important information above the fold, so it is visible without scrolling a page down.*

4.3 Ground Truth

For real data, we have adopted two methods to obtain it. One is to count the number of gazes of the volunteers browsing the webpages, and the other is to use questionnaires to allow each participant to rate the interest, proficiency, and difficulty of each webpage after browsing these 10 webpages. Of course, we also did a control experiment here. We also asked the same participants to perform the same operations on the computer and mobile pages, and then recorded two pieces of data.

Because we don't have an eye tracker, we adopted a manual counting method and mainly recorded three sets of data: the number of eye fixations on the text, the number of fixations on the non-text, and the average fixation duration. We manually count the number of fixations to fix the time, so there may be a little error in the data. We believe that if a participant looks at a text multiple times, it means that the text is difficult to understand; if a participant looks at non-text multiple times, it may indicate that he is very concerned about the main article; if a participant has a longer stay The process of reading a single word in a certain text is very difficult. The gaze on the main article-there is an article on each page-is counted as a text gaze; the gaze on the remaining pages is counted as a non-text gaze.

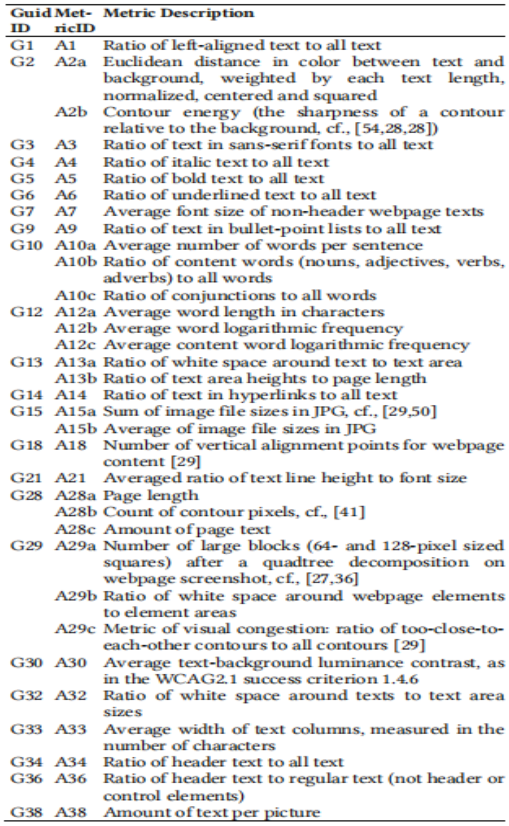
For the questionnaire survey, we mainly let each participant go to the page after browsing the page, and subjectively give a score to each one. Of course, this is also a controlled experiment. Each participant needs to rate both the computer-based and mobile-based pages. There are three main scoring indicators, one is the difficulty of reading the webpage, the second is their interest in the topic of the article, and the third is the scoring of the familiarity of the displayed website, because interest and familiarity may affect their reading Mode [1].

4.4 Manual evaluation

Manual evaluation mimics the real-world criteria people use in web page evaluation. We refer to the participants in this sub-study as experts to distinguish them from the eye-tracking sub-study participants, even if they have different levels of expertise. We recruited two experts, because the professionals I can know are limited, so I selected the best experts that I can find. Both of them are experienced people with Internet dyslexia. So I invited both of them to give a subjective report. Similarly, a control experiment was also done this time. The two of them need to rate these 10 pages on the computer and mobile phones in turn. I provided them with these 39 guides, and then waited for them to become familiar with the guides, and then rated the reading of these ten pages on different devices. This meeting took almost half an hour. We also found that the two experts gave similar scores for the same page, whether it was based on mobile or computer. This also shows that the display of the same page on different devices will not affect the readability, even if there are some small differences in the screen size and arrangement.

4.5 Automatic evaluation

We rely on the described web page readability features [2] and text complexity [4] measures to match some of the 39 criteria with automatic measures (Table 2).



Guid ID	Met-ricID	Metric Description
G1	A1	Ratio of left-aligned text to all text
G2	A2a	Euclidean distance in color between text and background, weighted by each text length, normalized, centered and squared
	A2b	Contour energy (the sharpness of a contour relative to the background, cf., [54,28,28])
G3	A3	Ratio of text in sans-serif fonts to all text
G4	A4	Ratio of italic text to all text
G5	A5	Ratio of bold text to all text
G6	A6	Ratio of underlined text to all text
G7	A7	Average font size of non-header webpage texts
G9	A9	Ratio of text in bullet-point lists to all text
G10	A10a	Average number of words per sentence
	A10b	Ratio of content words (nouns, adjectives, verbs, adverbs) to all words
	A10c	Ratio of conjunctions to all words
G12	A12a	Average word length in characters
	A12b	Average word logarithmic frequency
	A12c	Average content word logarithmic frequency
G13	A13a	Ratio of white space around text to text area
	A13b	Ratio of text area heights to page length
G14	A14	Ratio of text in hyperlinks to all text
G15	A15a	Sum of image file sizes in JPG, cf., [29,50]
	A15b	Average of image file sizes in JPG
G18	A18	Number of vertical alignment points for webpage content [29]
G21	A21	Averaged ratio of text line height to font size
G28	A28a	Page length
	A28b	Count of contour pixels, cf., [41]
	A28c	Amount of page text
G29	A29a	Number of large blocks (64- and 128-pixel sized squares) after a quadtree decomposition on webpage screenshot, cf., [27,36]
	A29b	Ratio of white space around webpage elements to element areas
	A29c	Metric of visual congestion: ratio of too-close-to-each-other contours to all contours [29]
G30	A30	Average text-background luminance contrast, as in the WCAG2.1 success criterion 1.4.6
G32	A32	Ratio of white space around texts to text area sizes
G33	A33	Average width of text columns, measured in the number of characters
G34	A34	Ratio of header text to all text
G36	A36	Ratio of header text to regular text (not header or control elements)
G38	A38	Amount of text per picture

5 Results

We reviewed the three pieces of data separately and then compared them against each other.

5.1 Ground Truth

The figure below shows the average fixation duration and number of fixations for the four participant groups based on the mobile web page, which shows that adults with dyslexia and children with dyslexia use fixation more often than ordinary children and adults. As expected, suffering from dyslexia will significantly increase the time and frequency of fixation. Children will increase the gaze time and the number of gazes on text, but the number of gazes on non-text does not increase. Suffering from dyslexia and childhood will further increase the fixed time. As a control group, based on the computer-based webpage, the data we got is similar to the data from the mobile-based webpage, and the situation is basically the same.

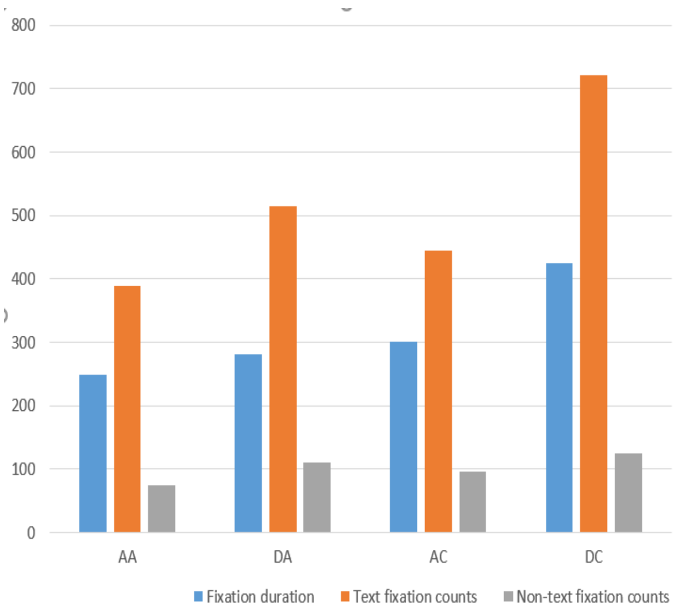


Figure 3. Means (SDs) of eye-tracking variables for averagereader adults (AA), average-reader children (AC), dyslexic adults (DA), and dyslexic children (DC)

5.2 Expert Evaluation

Although I only invited two experts, they both attached great importance to this experiment and took it very seriously. Their evaluation results show that most of the mobile-based website pages fully comply with the three criteria (G1, G4, G6), which means that our data set does not have enough variance to test these criteria; we start from further analysis Excluded them.

We assume that the average of the two expert scores for each webpage of each guide is closer to the actual score of the webpage and the guide than a single expert's score, and estimate the expert's ability to comply with the guidelines as the average of the difference between the average and the expert's score . The ability to persist is not related to the experience of dyslexia. However, dyslexia experts tend to use more extreme scores. The more experienced people will not reduce the evaluation time: the evaluation time of each page time of the experts has nothing to do with their design experience.

5.3 Automatic Evaluation

After scaling and filtering the calculated data, we looked at the histogram of the calculated variables, which showed that several metrics have only limited variance. Whether it is the control group or the experimental group, the data we get are similar.

5.4 Expert and Automatic Evaluation and Ground Truth

To test whether expert evaluation can be replaced by automatic evaluation, or whether expert evaluation captures readability differences that automatic evaluation cannot achieve, we explored such criteria in a series of mixed linear models. We entered the expert score and the algorithm score separately in the model, and entered the two together to provide three models for each criterion-and checked whether the model performance has improved. If adding algorithm scores as predictive indicators in addition to expert scores does improve performance, we believe that expert evaluation can be replaced by automatic evaluation. If both the algorithm and the human score contribute significantly to the performance of the entire model, then we believe that automatic and manual evaluation are complementary.

Through experiments, we found that some criteria can be automated or partially automated, but humans need to recheck the results of the automated evaluation. This group contains some text complexity criteria that the algorithm can measure, or text format criteria that include clauses that define its scope, which may be problematic for algorithm processing. Such guidelines include: the brightness contrast guide (G30), which has a clause that defines its scope as non-decorative elements; the guide on hyperlinks (G14), which stipulates that the hyperlinked page should be related to the main content, that is, a single column Guidelines (G18), which specify that only important main content should be in the same column, and heading guidelines (G34), the algorithm indirectly evaluates in the context of the web page by counting the amount of header text and assuming that these texts are relevant.

Finally, some criteria are difficult to automate, and human evaluators need to use them without the help of algorithms. The last set contains guidelines that need to understand and explain the topic of the web page, and these algorithms will be difficult to complete. Such guidelines include the use of jargon (G14), symbolic icons for navigation elements (G22), the main points of sentences at the top (G24), and labels and menu items to be concise (G35). These guidelines will require explanations of what content is suitable for a particular audience, if an icon meaningfully describes an item, what are the main points of sentences and paragraphs, and whether menu items actually link to the topic of the page.

6 Discussion

The main research of this work is about the similarities and differences between readability assessment on mobile pages and computer pages. When evaluating web pages according to specific criteria, the performance of automatic evaluation on human web pages is better than expert evaluation, especially those algorithms related to web page legibility and low-level functions of text format. However, the judgment of multiple standards still requires manual evaluation and understanding and interpretation of Web content.

Based on this, we can conclude that on mobile devices, automatic evaluation cannot completely replace manual evaluation, but can only be used as a supplement to manual evaluation. For some lengthy and single criteria, automatic evaluation can be used as the evaluation method, and the personnel can recheck the automatic evaluation. For some criteria that are difficult to automate, such as those that need to understand and explain web topics, human evaluators need to use them without the help of algorithms.

In terms of differences, it is mainly reflected in the increase in the amount of information obtained by experts at the same time due to the smaller screen on mobile devices. Compared with the web version, expert evaluation on mobile devices takes longer, that is, the efficiency of manual evaluation. Lower. On the other hand, the reading criteria of mobile devices are more difficult to automate, that is, the corresponding scope of automatic evaluation is smaller. Therefore, a more reasonable evaluation method is partial automation. Some evaluation criteria that are difficult to measure directly. For example, such guidelines include: Brightness Contrast Guidelines (G30), which has a clause that defines its scope as non-decorative elements; Hyperlink Guidelines (G14), which stipulates that hyperlinked pages should be The main content is related, that is, the single-column rule (G18), which specifies that only important main content should be in the same column, and the heading rule (G34), the algorithm calculates the number of heading texts and assumes that these texts are relevant, in the web page Indirect evaluation in the context of

The results show that evaluators with less evaluation experience are more likely to benefit from relying on the guide list, and at the same time more adaptable to mobile devices: on mobile platforms, we have observed that novices have the problem of lower observation efficiency compared to veterans. However, with the help of the algorithm, the observation efficiency of both is within an acceptable range; at the same time, novices have made faster progress on the mobile platform. After multiple evaluations of similar evaluation content, the evaluation efficiency of the two gradually Convergence. But for criteria without the help of algorithms, evaluation experience is still the main influencing factor.

7 References

[1] Schiefele, U., Krapp, A. Topic interest and free recall of expository text. Learning and individual differences, 8, 2 (1996), 141-160.

[2] Miniukovich, A., De Angeli, A., Sulpizio, S., and Venuti, P. Design Guidelines for Web Readability. In the 2017 Conference on Designing Interactive Systems (2017), ACM, 285-296

[3] Graesser, A. C., McNamara, D. S., Louwerse, M. M., and Cai, Z. Coh-Metrix: Analysis of text on cohesion and language. Behavior research methods, instruments, computers, 36, 2 (2004), 193-202.

[4] Bates, D., Maechler, M., Bolker, B. Walker, S. Fitting linear mixed-e ects models using lme4. Journal of Statistical Software, 67 (2014), 1–48.

[5] Arditi, A. and Cho, J. Letter case and text legibility in normal and low vision. Vision research, 47, 19 (2007), 2499-2505.

[6] Aziz, F. A. and Husni, H. Interaction Design for Dyslexic Children Reading Application: A Guideline. In Knowledge Management International Conference (KMICe) (2012), 682-686.

[7] Brown, C. M. Human-computer interface design guidelines. Intellect Books, 1999.

[8] Cassim, R., Talcott, J. B., and Moores, E. Adults with dyslexia demonstrate large effects of crowding and detrimental effects of distractors in a visual tilt discrimination task. PloS one, 9, 9 (2014).