Design Guidelines for Web Readability

Aliaksei Miniukovich[*]**, Antonella De Angeli** [*,#]**, Simone Sulpizio**[$]**, Paola Venuti**[^]

[*] DISI, University of Trento, Italy

[#] School of Computer Science, University of Lincoln, UK

[$] Faculty of Psychology, Vita-Salute San Raffaele University, Milan, Italy

[^] DiPSCo, University of Trento, Italy

{aliaksei.miniukovich, antonella.deangeli, paola.venuti}@unitn.it, sulpizio.simone@hsr.it

ABSTRACT

Reading is fundamental to interactive-system use, but around 800 million of people might struggle with it due to literacy difficulties. Few websites are designed for high readability, as readability remains an underinvestigated facet of User Experience. Existing readability guidelines have multiple issues: they are too many or too generic, poorly worded, and often lack cognitive grounding. This paper developed a set of 61 readability guidelines in a series of workshops with design and dyslexia experts. A user study with dyslexic and average readers further narrowed the 61-guideline set down to a core set of 12 guidelines – an acceptably small set to keep in mind while designing. The core-set guidelines address reformatting – such as using larger fonts and narrower content columns, or avoiding underlining and italics – and may well aply to the interactive system other than websites.

Author Keywords

Web; readability; accessibility; dyslexia; WCAG 2.0.

ACM Classification Keywords

H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces – Style guides.

INTRODUCTION

Reading is a major part of almost every task on the Web. High readability makes the tasks substantially easier, which reduces users' frustration and time wasting. High Web readability and its positive outcomes may be a luxury for the regular users, but become almost a necessity for the users with reading difficulties. For these users, reading and understanding text requires more time and effort than for regular users. Unless website designers are ready to neglect them – which many government and non-profit websites are forbidden to do – they should design for high readability.

Dyslexics are a notable group among the users with reading difficulties, as they – unlike, for example, low-literacy users – have no problem comprehending complex concepts, only

decoding concept presentation. This quality of dyslexics might help researchers disentangle poor concept presentation from genuinely high concept complexity: re-design would address presentation and design practice would want to have recommendations about presentation.

Psychology has accumulated a formidable amount of theoretical knowledge on dyslexia and readability, but little of it has reached design practice. Abstract models and theoretical findings cannot be directly applied in design practice, and intermediate-level tools [39] – such as design guidelines – are needed to bridge theory and design practice. However, HCI research has given insufficient attention to the development of readability guidelines: They still require much improvement before they are finalized and taught to future Web designers: many of them lack clarity or empirical support, contradict each other, or are poorly formulated.

This paper adopted a systematic approach for developing an actionable set of Web readability guidelines. We pooled, reviewed and reduced a large number of existing guidelines in a series of workshops with dyslexia experts and designers. This activity resulted in a set of 61 non-composite, concise, and actionable guidelines based on psychological evidence. Applying all 61 guidelines, however, might take too much time and effort for some projects, so we attempted to define a preliminary sub-set of core guidelines. For this purpose, we triangulated three datasets: (a) webpage readability evaluation by 13 dyslexic and 23 average readers, (b) expert opinion of 14 trained evaluators (designers and dyslexia experts); and (c) automatic evaluation of webpage compliance with the WCAG 2.0 (Web Content Accessibility Guidelines) [33] and selected readability guidelines. The study highlighted 12 guidelines that affected reading for either dyslexic or non-dyslexic users, or for both of them.

In the remainder of the paper, we review the readability and dyslexia related literature, and list and analyze a big set of readability guidelines. We then describe guideline selection in the workshops and user study. The paper concludes with several implications for future research and design.

RELATED WORK

Dyslexia affects a large group of users, up to 10% of population [10], and have been addressed in both psychology and HCI.

Dyslexia

Individuals with dyslexia have difficulties in accurate or fluent word recognition, poor spelling and decoding abilities; at the same time, these individuals have adequate instruction, intelligence, and sensory abilities [3]. Several typical errors that dyslexics tend to make include letter inversion (e.g. b for d, p for q) or reversal (e.g., m for w), letter transposition (e.g., felt and left), word reversal (e.g., tip for pit). Apart from these specific errors, individuals with dyslexia generally struggle to recognize, spell and decode written words. Reading becomes more difficult, which further undermines text comprehension.

Although cross-linguistic differences exist in terms of easiness of learning to read – which is easier in transparent languages, such as Italian where the letters (graphemes) are in most instances uniquely mapped to each of the speech sounds (phonemes), than in non-transparent languages where this mapping is more ambiguous (like English or French, [27]) – it seems that developmental dyslexia has a similar rate of occurrence across languages [10], and has universal basis in the brain that are independent of the spoken language [14].

A prominent theory of dyslexia etiology argues that developmental dyslexia is caused by a deficit in visuo-spatial attention and originates in the mechanisms responsible for the attentional shifting [31]. Issues with attentional shifting may then result in crowding [8,34] – a perceptual phenomenon occurring when an object becomes more difficult to detect or discriminate from its surroundings, if it is presented in isolation. Compared to typical readers, individuals with developmental dyslexia are abnormally affected by crowding [32,12,28].

Readability Guidelines

Past research advocated using design guidelines instead of other means to tackle Web accessibility issues [7]. The guidelines concisely summarize good design practices, fit well in design classes, and thus, can increase awareness of accessibility solutions. Such awareness would address accessibility issues in a more robust fashion than, for example, specialized accessibility tools [35,25] Furthermore, guidelines also decrease reliance on specific technological platforms.

However, most of the accessibility guidelines addressing reading focused on the need of blind [29,16] or elderly users [11], with scant attention given to dyslexic readers. Other studies focused on accessibility in general [7,33], which make it difficult to understand what guideline helps which category of users. However, some specific set of guidelines targeting dyslexic users exist. For example, Santana et al. [24] surveyed web accessibility and dyslexia, proposing a set of guidelines set aimed at developing dyslexia-friendly websites. Jackson [9] assembled a similar guideline set, focusing on typographic issues (e.g., font size, inter-letter space) that could help dyslexic readers.

Yet, building a comprehensive set of readability guidelines remains a challenge, mostly because past studies tested only few guidelines at a time and are hard to compare due to their methodological heterogeneity (e.g., [2,6,37,20]). Further, different guideline sets include recommendations that contradict each other, e.g., the guideline *Line spacing of 1.5 is preferable* was supported in [20,37], but not in [22,26]. In other cases, a guideline may contradict empirical findings from psychological studies. For example, *Avoid capital letters* [1,5] contradicts the finding of dyslexic children recognizing words in upper-case better than in lower-case [15]. Finally, some guideline sets included recommendations lacking empirical validation, as in the case of *Avoid splitting words at the end of a sentence, i.e. do not use hyphens*, or *Use bullet points and numbering rather than continuous prose* [18,1].

GUIDELINE REVIEW

We reviewed existing readability guidelines in four rounds. The first round assembled an initial guideline set; the other three rounds – initial guideline analysis, workshop with dyslexia experts, and workshop with user interface experts – reviewed and iteratively excluded guidelines from the initial set. We considered only the guidelines related to both readability – the effort spent on reading a text – and dyslexia. Our review aimed at assembling a manageable-size set of concise and potentially useful guidelines to be tested in a user study.

Guideline Collection

In November 2015, we queried Google Scholar for the papers mentioning "dyslexia" and "guidelines" and searched through the 50 first pages of the search output. We further looked through the references of the retrieved papers for other possible sources of guidelines for dyslexia. In this way, we collected a total of 10 documents with guidelines on how to design dyslexia-friendly web or book pages [1,5,24,4,9,13,18,20,30,36,19]. The sets of guidelines partially overlapped. Some recommendations appeared in most of the documents (e.g., *Use left-justified with ragged right edge*, [18,24,4,1,30]) while other guidelines were unique (e.g., *Use single space after period* [30]). We removed the duplicates, retaining the more concisely formulated one. After pooling the guidelines from all the documents and filtering out duplicates, we ended up with 138 potentially helpful guidelines. We also considered 61 W3C-recommended WCAG 2.0 guidelines [33], as they are widely perceived as the standard set of guidelines about accessibility issues, including dyslexia. The WCAG set was treated separately as we wanted to look at the overlap between the dyslexia-related and WCAG guideline sets. This would let us conclude how well dyslexia is addressed in the WCAG 2.0 guidelines.

Guideline Analysis

Two authors reviewed the entire corpus of guidelines and excluded those belonging to one of three categories. The first category included the text-to-speech webpage adaptation guidelines (e.g., "*Use the <abbr> tag to explain*

abbreviations and acronyms" [18] helps the text-to-speech software parse the website, but does not improve the readability of website). The second category included dyslexia-unrelated guidelines (e.g., "*Avoid green and red/pink combinations as these can cause confusion for colour-blind users*" [18]). The third category included guidelines with an extremely broad and unfocused scope (e.g., "*Colour preferences vary*" [5]). During the coding, the two authors independently analyzed all guidelines, flagged those belonging to one of the three categories, and discussed eventual coding discrepancy till a consensus was reached.

Many of the guidelines included composite statements: they reported several independent pieces of information, each of which could be considered a guideline on its own. During the analysis, we split these composite guidelines in simple, one-instruction recommendations. For example, the guideline "*Consider short paragraphs. Avoid complicated language/jargon. Use short, simple, and direct sentences…*" [24] was divided into three separate recommendations. Many of the guidelines were then re-worded to make them more actionable and concise. For example, "*Very large graphics make pages harder to read*" [5] became *Avoid very large graphics that make pages harder to read*; whereas "*Whenever possible, structure index pages considering a logical order involving tasks sequence or structure (e.g., when involving information that can be structured as part-whole), without requiring exclusively the alphabetical ordering*" [24] became *Arrange the items on index pages in a logical order (e.g., by function); avoid exclusively using the alphabetical order.* The total number of guidelines became 98, with 17 of them extracted from WCAG 2.0. Overall, we noticed the WCAG 2.0 predominantly addressed website use and functionality; text structure and appearance were largely ignored.

Dyslexia Experts

The first workshop addressed three issues. First, 12 guidelines about five design aspects contradicted each other (e.g., Rainger [18] called for using bold text for highlighting keywords and concepts, whereas [1] advocated against having bold anywhere on a page). Second, some guidelines lacked empirical evidence directly supporting their validity, though they could be inferred from the expert-level knowledge of dyslexia. Third, the connection of selected WCAG guidelines to dyslexia remained to be checked.

Method

Three experts in dyslexia – two practicing psychologists and one academic – were invited to resolve contradictions and review the validity of guidelines. Participants were invited in a quite room, introduced to the study and handed over the guidelines. Each guideline was printed on a card; the cards also included an id for easier identification. The WCAG-extracted guidelines were printed on yellow paper, while the other guidelines were printed on white paper, so

the experts could clearly differentiate the two sets. We first asked the experts to individually review contradicting guidelines, and then to discuss and select the guidelines that they thought had more supporting empirical evidence. We then asked them to link the Dyslexia guidelines to the WCAG guidelines that addressed the same reading-related phenomena. Such analysis would check if the selected WCAG guidelines were indeed related to dyslexia. Finally, experts rated each guideline considering "*How much the complete violation of a guideline will degrade webpage readability for dyslexics, relative to the full adherence to the guideline*", using a 7-point Likert scale (from *not at all* to *completely*).

Results

Table 1 presents the results of dyslexia experts resolving the contradictions among the guidelines. One guideline was chosen and retained for each of five design aspects, which solved the first issue. To solve the second issue, experts' scores of guideline impact were converted into z-scores and averaged per each guideline. The total of 12 guidelines with average scores below -1 (i.e., with the impact more than one standard deviation below the average) were excluded from further analysis, as they were deemed unlikely to have an impact on readability. Addressing the third issue, the experts linked 16 of selected 17 WCAG-based guidelines to between 21 and 33 dyslexia guidelines (the experts did differ in the criteria they used to count two guidelines as linked; the workshop lasted for over two hours and we could not delay the experts for any longer to reach consensus). Finally, based on experts' feedback and additional analysis, we further excluded several guidelines that the experts failed to find meaningful (e.g., the advice of avoiding digressive hyperlinks [18]). WCAG-based and dyslexia guidelines were merged, with 67 guidelines remaining in total.

Design aspect	Guideline
Line Spacing	- *Line spacing of 1.5 point is preferable.*
Color combination	- *Use an off-white color for your background, like light gray or tan; use dark gray for text instead of pure black.*
Bold	- *Use bolding to highlight in order to emphasize keywords and concepts.*
Underline	- *Avoid underlining large blocks of text as it makes reading harder.*
Font Size	- *Use a minimum of text size 12pt or 14pt.*

Table 1. Webpage aspects and corresponding guidelines that dyslexia experts chose as the most effective.

Design Experts

The second workshop revised the guidelines with web designers, who the guidelines are ultimately aimed for. The workshop intended to address three issues: to check the clarity of guidelines, to exclude the guidelines that were less-relevant for webpages, and to categorize the guidelines in groups.

Method
Three experts with working and/or teaching experience in web design or development participated in the workshop. All 67 guidelines were printed on same-size paper cards and placed on a large empty desk in three groups, one for each expert (Figure 1). The front of each card featured a number and the guideline text. The back was empty. We introduced the experts to the study, asked them to look through the guidelines and indicate those that required further explanation or re-wording. This addressed the issues of clarity and relevenace. To address the issue of categorization, we administered two rounds of single-criterion card sorting [23], with a few adaptations. In the first round, the experts were given no pre-defined sorting criteria; they defined themselves the sorting criteria they would adopt and the resulting categories, and finally sorted the guidelines. In the second round, the experts were given three pre-defined sorting categories – *appearance*, *content*, and *functionality*, cf [4] – and also instructed to use the *other*, *not sure* or *not applicable* categories if needed. After individually sorting the guidelines in the pre-defined categories, the experts discussed and tried to resolve eventual differences in sorting.

Figure 1. Card sorting during the workshop with web experts. Each expert was assigned to a desk with a set of 67 guideline cards; yellow masking tape was used to split each desk in sections; each section provided space for one card group.

Results
Six guidelines were excluded from further analyses because experts indicated them as duplicates (e.g., *Avoid using long and dense paragraphs* and *Consider using short paragraphs*), as inapplicable to a general webpage (e.g., *Use flow charts to explain procedures and diagrams to explain spatial concepts*), or as overly vague (e.g., *Consider that people may not be able to distinguish background from foreground in images and may be distracted*). The number of guidelines in the set was thus reduced to 61.

In the first round of card sorting, each of the three experts suggested six categories for the guidelines. Little consistency was present in these categorizations, with 16 unique category labels used. Instead of choosing one of categorizations and discarding others, we tried to detect a configuration that would rely on all three experts' categorizations. We created 18 (3 experts * 6 categories each = 18) dummy variables for each guideline. The

possible variable values were one (a guideline was in a category) and zeros (not in the category). We then calculated the L_1 pairwise distances among the guidelines and used the distances to conduct a hierarchical cluster analysis. The resulting tree-like hierarchy of guidelines was cut in n clusters. The cutting was repeated nine times, for n = 3 to 12 clusters. Each time we reviewed the resulting clusters for the consistency of their member guidelines. The five-cluster configuration seemed the easiest to interpret, and we retained it (Table 2, 1st column). The emerged five clusters emphasize the important aspects of design, and could be treated as the general readability dimensions in future studies and for readability-guideline presentation.

In the second round of card sorting, the experts individually sorted all guidelines in the three pre-defined categories (none of the *other*, *not sure* or *not applicable* categories were used). When later asked to compare their categories and resolve differences, the experts could not agree on a single categorization in a reasonable amount of time, arguing about almost a half of all guidelines. For example, each of them assigned *Limit the amount of content on a page to avoid scrolling* to a different category and they all had a strong justification for their choice as they claimed that the guideline could influence any of the webpage *appearance*, *content*, and *functionality*. These three categories appeared to semantically overlap, which was sub-optimal and we dropped the pre-defined category data from further analyses.

USER STUDY
The three rounds of review reduced the initial set of 199 guidelines down to 61 guidelines, which were still too many for the use in design practice. To further reduce the set down to a few core guidelines, we conducted a study, collecting three types of data: ground-truth readability data by dyslexics and average readers, compliance with guideline data by experts, and automatic readability evaluation data.

Study Sample
We opted for sampling and studying *multiple* real-life webpages. The other, classical-psychology approach would be sampling *few* webpages and carefully manipulating their appearance to a possibly unrealistic extent. Such approach would allow for definitively validating the guidelines, but would be sub-optimal for the main goal of the study: finding the core guidelines that applied to most of realistic webpages.

The sample included 36 full-page screenshots (1185 pixel wide) of Italian webpages from three domains, *local government*, *health*, and *education*. The websites in such domains should comply with accessibility standards: government and education websites were subject to the disability-discrimination laws, whereas health websites often specifically targeted people with disability, including reading disabilities.

Category	Guideline	Scope	Excluded
Visual appearance	Use left-justified text with ragged right edge. [1,30,5,18,4,24]	A	
	Use an off-white color for your background, like light gray or tan; use dark gray for text instead of pure black. [24,1,18,20,4]	A	
	Use a plain, evenly spaced sans serif font such as Arial and Comic Sans. [18,30,5,1,24,4,9]	A	
	Avoid using italics in the main body of the text. [30,5,1,24,4,9]	A	Lr
	Use bolding to highlight in order to emphasize keywords and concepts. [5,4]	A	
	Avoid underlining large blocks of text as it makes reading harder. [18,30,5,1,4]	A	Lr
	Avoid using more than one whitespace after period. [30]	A	Lr
	Use a minimum of text size 12pt or 14pt. [18,5,1,24,36,20]	A	
	Avoid capital letters, apart from the beginning of sentences, abbreviations, and where it is grammatically correct. [18,5,1,24,20]	A	
	Ensure navigation menus use a text size that allows for comfortable reading. [24]	E	
	Whenever text highlighting is needed, consider using boxes, border, and background color to call user attention. [18,24,5]	M	Na
	Keep the between-line spacing of 1.5 point. [5,18,1,24,4]	A	
	In the body of text, always underline hyperlinks as users expect to be able to recognise hyperlinked text or images. [18,24]	A	Na
	Avoid the fonts in which letters like b-d or p-q are perfectly mirrored letters. [4]	A	
	Ensure navigation menus differ visually from the main body of webpage. [24]	E	
	Avoid fonts with unusually shaped letters. [18]	A	
	Avoid splitting words at the end of a sentence, i.e. do not use hyphens. [18,1]	A	Lr
	Use enough white space between webpage elements. [5,18,24,36]	A	
	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. [18,5,24,36,33]	A	
	Ensure webpage elements (buttons, links, icons, etc.) that have the same function also have the same look. [33]	E	
	Keep the white space between paragraphs of at least 1.5 times the space between text lines. [33,24,20]	A	
	Avoid formatting texts in large-width columns, especially Asian logogram texts. [18,33,24,20]	A	
	Ensure visited links are visually distinct from unvisited links. Visited links are often given in a purple colour. [24,5,18]	W	
	Avoid using blinking or moving text. [24,18,5,13]	W	
	Avoid dynamic menus or menus that use transparency, because users with dyslexia may find hard to contrast them with the background partially visible. [24]	W	
Structure of content	If appropriate, use bullets or numbers rather than continuous prose. [18,5,1,24,20]	M	
	Use graphics, images, and pictures to break up text, whilst remembering that graphics should be relevant to the material and not distract from the content. [18,30,24,36,4]	M	
	Use short, simple sentences in a direct style. [18,30,5,24]	M	
	Use active rather than passive voice. [5,1]	M	
	Avoid complex language and jargon. [1,24,36]	M	
	Consider using short paragraphs. [18,30,24,5]	M	
	Avoid images that are 'busy', cluttered, include irrelevant or too much extra detail. [1]	M	
	Avoid placing images above text or text around images. [1]	A	
	Place the main point at the very top of page. Place any other important information above the fold, to minimize the risk of users losing their place after scrolling. [5,13]	E	
	Place important content in a single main column, so users don't have to scan the page and pick out design elements in a two-dimensional layout. [13]	M	
	Be conscious of where sentences begin on the page. Starting a new sentence at the end of a line makes it harder to follow. [5,18,20]	M	
	Provide clear intuitive categories for menus and links. [18,24,36]	E	
	Put the main point of sentence or paragraph into the beginning of the sentence or paragraph. [1]	M	
	Always put the search box in a clear obvious position, usually the top of the page. [18]	E	
	Limit the amount of content on a page to avoid scrolling. [18,24,13]	E	
	Ensure Web pages have titles that describe their topic or purpose. [18]	E	

	Scope	Excluded
Ensure the purpose of each link can be determined from the link text alone, without reading much text before the link. [24,36,33]	M	
Ensure headings and labels concisely describe the topic or purpose of page sections and elements. [18,33]	M	
Use section headings to organize the content. [18,33]	M	Na
Arrange the items on index pages in a logical order (e.g., by function); avoid exclusively using the alphabetical order. [24]	W	
When building a site index, use visualizations revealing hierarchy of website pages. [24]	W	
Provide a printer-friendly version of webpage. [18]	W	
Let the user change text and background colors, and enlarge text. [33]	W	
Provide users enough time to read and use content; avoid setting time limits on reading. [33]	W	
Navigation		
Embed in Webpage texts the hyperlinks to the pages with the text-related concepts. [24,36]	M	
Use text and symbolism for navigational elements that are truly representational or a well-known concept e.g. a house for home. [18]	M	
Use a breadcrumb trail (e.g., "Home page > section 1 > sub-section 1.1") to let the user understand their location on a website. [18]	E	
Ensure navigation menus group information by function. [18,24,36]	E	
Avoiding nested menus (e.g., containing items, sub items, sub-sub items, etc.); opt for broader one-level menus. [13]	W	
Allow the user to go back and re-read information. [4,33]	W	
Functionality		
Include in your website common "helper" pages, such as homepage, contact us, site map, references and links, copyright and disclaimer, and accessibility statement pages. [18,5,24]	E	
Offer internal search on a website; provide self-completion, and orthographic verification to point errors and ease correction. [5,18,24]	W	
Avoid horizontal scrolling on webpages. [33]	W	
Ensure navigation buttons and links that are repeated on multiple pages of a website occur in the same relative order and similar position across the webpages. [33]	W	
Communication		
Use the lists of dos and don'ts, which are more useful than continuous text to highlight aspects of good practice. [18]	M	Na
Provide a version of regular texts, which doesn't require reading ability more advanced than the lower secondary education level. [33]	W	

Table 2. The list of guidelines after 3 rounds of review. Guideline category was automatically derived from the expert-based categorization. *Scope* describes the area of guideline effect: W – applies to entire website; E – applies to an entire webpage; A – applies to all webpage elements; M – applies to the main body of content. *Excluded* describes the guidelines excluded during statistical analyses due to limited range (Lr) or non-applicability (Na).

The selected webpages featured one of six types of information: *rules & regulations, service being offered, piece of news, public announcement or public event, health-related article,* and *educational course information*. The webpages were unrelated to the region our participants were based in, which should have reduced familiarity biases. All webpages had to feature a complete article requiring no specific expertise to comprehend.

The authors collected an initial large sample and reduced it to satisfy a set of requirements. We strived to represent well the present-day Web and ensured retaining both poorly and well-designed webpages. The webpages that looked broken or abandoned were omitted from the sample. The webpages with an excessive amount of text were also omitted – too much text to read could needlessly burden our participants. Finally, our sample did not include the webpages featuring too much advertisement or dubious content, such as alternative teaching techniques or alternative medicine.

Data Collection
Data were collected in three separate rounds. First, we collected the baseline readability scores for each webpage:

dyslexics and average readers read the webpages and rated their readability. Second, dyslexia- or design-experienced reviewers rated the webpages on their compliance with the subset of 47 out of 61 guidelines. Lastly, the expert-based evaluation was complemented with automatic measures: we created automatic measures for several guidelines and computed measure estimates.

User Ratings
An online user study collected the baseline ratings of webpage readability. We emailed the link to the study website to a local dyslexia association that we collaborated with in a number of research projects. The greeting webpage contained a study description and instructions (e.g., how to maximize a browser window), and asked for participants' consent to the use of their data.

Participants. A total of 36 people completed the study; 13 of them reported having been certified with dyslexia (7 female) whereas 23 (11 female, 1 refused to specify a gender) did not. We call these two groups as dyslexics and average readers. Dyslexics reported spending 4.65h a week on reading; average readers reported spending 6.20h a week

on reading. Dyslexic users were slightly younger than average readers: no dyslexics were older than 40y, whereas five out of 23 average readers were (we collected the age data using age brackets).

Procedure. After reading the instructions and agreeing to their data being used in research, each participant read and rated 10 webpages using a 7-point semantic differential scale ("*reading this webpage was very simple/very difficult*"), cf., [17]. This measure has been successfully used in past psychological research [38, 40, 41]. To ensure that participants carefully read the pages, we asked three content-related questions: two multiple-choice questions ("*Choose a category that describes the website genre the best*" and "*Choose a category that describes the webpage content the best*") and one open question ("*Describe the webpage info using up to 5 keywords*"). At the end of the study, participants filled in a brief demographic questionnaire. For each webpage, we recorded two types of data: readability rating and time to read a webpage (time between a webpage appearing on the screen and user clicking a 'proceed' button).

Expert Ratings
To evaluate the webpages on their conformity to guidelines, we conducted a study with designers and dyslexia experts.

Participants. We sought participants that could fluently speak both Italian and English, since the evaluated webpages were in Italian and the guidelines were in English. We also required participants to be experienced in either web development or dyslexia, or in both. In total, 14 experts (5 female; mean age = 29.9y, SD = 5.1y) working for or with the university (graphic designers, developers, and research personnel) participated in the study; their reported design and dyslexia expertise ranged from basic to substantial, which represented the whole spectrum of people who might need to apply the guidelines in practice.

Guidelines. Out of the total of 61 guidelines that were previously selected, 14 guidelines (Table 2, column *Scope* values of W – **w**ebsite) could not be tested on screenshots, as they would require fully functional websites as stimuli. The remaining 47 guidelines (Table 2, column *Scope* values of E, M and A) were applicable to entire webpages (E), main body of text (M) or all webpage elements (A), and could and were tested.

Procedure. The experts participated in the study individually. After reading instructions and agreeing to the use of their data, each expert rated seven webpages on their compliance with the 47 guidelines. A webpage screenshot was shown on the left of the screen; the guidelines were reported on the right (Figure 2). We used a 7-point Likert scale, with the question "*the webpage complies with the guideline* [guideline text]" and anchors "*not at all*" and "*completely*". For each question, the experts could also select on of two extra answers, "*The guideline doesn't apply to the webpage*" and "*I don't understand the guideline*".

The guidelines were presented in 3 clusters based on their scope: applicable to the entire webpage as a unit, applicable to the main body of content, and applicable to all individual visual or textual elements of webpage. Each group was accompanied by a short instruction on how to apply the guidelines. Experts had no time limit to complete the experiment, which on average lasted 85 minutes. The experts did not report fatigue. For each guideline, we recorded three types of data: compliance rating (1 to 7), if applicable to a particular webpage (binary), and if understood (binary).
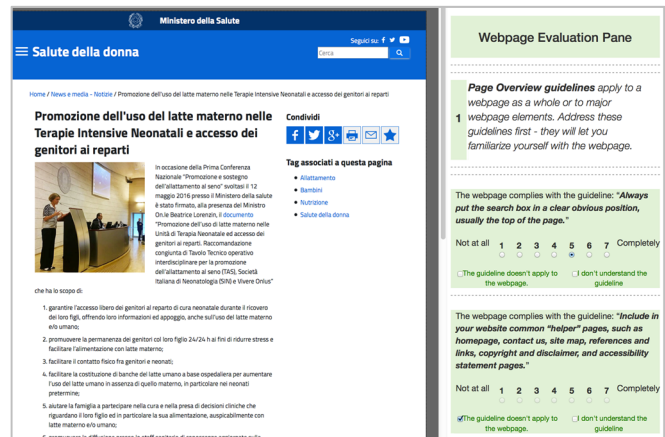

Figure 2. The set-up of expert evaluation of webpage compliance with guidelines.

Automatic Measures
The compliance of each web page with the WCAG 2.0 accessibility guidelines was automatically measured. First, we queried the API of online WCAG 2.0 validator *Achecker*[1] and recorded the number of accessibility *errors*, and *likely errors*. Each webpage size in bytes and in number of HTML elements was also recorded. We expected the numbers of errors to depend on both webpage accessibility and webpage size (larger webpages had more room for error) and needed to normalize error numbers by page sizes, i.e., we would calculate the number of errors per unit of page size.

In addition, we operationalized several of the guidelines to be automatically measured (Table 3). One might have expected algorithms to evaluate low-level aspects of webpages (e.g., the amount of italics on a page) better than human experts could: expert-based evaluation should have been supported with algorithmic evaluation. To compute the measures for the guidelines, we implemented a plugin for the Mozilla Firefox v.30 browser. This solution allowed accessing webpage content exactly how it looked in the browser, after all CSS were applied and JS scripts executed on the webpage. The plugin also let us exclude from the analyses all invisible webpage parts.

---

[1] http://achecker.ca/checker/index.php

Page aspect	Measure	Guideline
Amount of text	The number of characters of visible text.	- *Limit the amount of content on a page to avoid scrolling.*
Headers	The proportion of text taken by headers.	- *Use section headings to organize the content.*
Text style	The proportion of text formatted as italic, bold, or underlined.	- *Avoid using italics in the main body of the text.* - *Use bolding to highlight in order to emphasize keywords and concepts.* - *Avoid underlining large blocks of text as it makes reading harder.*
Font size	For each font size, we counted the number of characters, weighed by font size, and normalized by the number of all characters.	- *Use a minimum of text size 12pt or 14pt.*
Font type	The proportion of text styled with serif, sans serif, cursive, and fantasy fonts.	- *Use a plain, evenly spaced sans serif font such as Arial and Comic Sans.*
Line spacing	For each piece of text, we calculated the height of text element minus the height of text itself and normalized by the height of text itself. The number of characters in each piece of text was weighed by the calculated index.	- *Keep the between-line spacing of 1.5 point.* - *Keep the white space between paragraphs of at least 1.5 times the space between text lines.*
Width of text column	For each piece of text, we calculated the number of characters that fit in one line. The size of each piece of text was then weighed by the column width, summarized, and normalized by the number of all page characters.	- *Avoid formatting texts in large-width columns, especially Asian logogram texts.*
Text alignment	The proportion of webpage text that was styled as left-aligned.	- *Use left-justified text with ragged right edge.*
Luminance contrast	For each piece of webpage text, we calculated the text-background luminance contrast. The length of text pieces was weighed by the contrast, summarized, and normalized by the length of all page texts.	- *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.*
Color difference	For each piece of webpage text, we calculated the Euclidian distances between text color and background color. The number of characters in the text pieces was weighed – large and small color distances received smaller weight; medium distances received bigger weight – summarized, and normalized by the length of all page texts.	- *Use an off-white color for your background, like light gray or tan; use dark gray for text instead of pure black.*

Table 3. The automatic measures for several readability guidelines.

Table 3 lists the implemented automatic measures and associated guidelines. All measures (except Text Amount) returned a single numeric value ranging from 0 to 1, which would indicate *how much* a webpage complied with a guideline. Out of many different elements on the same webpage (e.g., headers, buttons or paragraphs), some elements complied with a guideline; others did not. The effect of all elements needed to be aggregated in a single score per webpage, which made only meaningful answering how much a webpage complied with the guideline.

Study Results
The study collected three types of data: baseline readability scores, experts' scores of webpage conformity to guidelines, and computed scores for some of the guidelines.

User Data Review
We reviewed readability data for irregularities and excluded the data of three average readers because they used only a narrow sub-scale (e.g., one participant gave 2 to all webpages) or strongly deviated from other participants

(e.g., the mean of one participant ratings was 6.3, more than double the average of all others).

The review of the expert-based data suggested that they could understand all guidelines, with a minor issue with "*Place the main point at the very top of page. Place any other important information above the fold, to minimize the risk of users losing their place after scrolling*" and "*Avoid formatting texts in large-width columns, especially Asian logogram texts*". Two experts found these guidelines difficult to understand, but we still kept them since the other 12 did not report any problem. We then reviewed the guidelines that the experts indicated as non-applicable to a webpage. Four guidelines that applied to less than a half of webpages were excluded from further analyses (Table 2, column *Excluded*). Finally, we reviewed the score histograms for each guideline and excluded four more guidelines due to their scores varying in a very limited range (e.g., mostly 7 and some 6). Such limited variance

indicated that all webpages fully followed a guideline, rendering the guideline not relevant for typical webpages.

Readability Validation

Dyslexics' readability ratings moderately correlated with non-dyslexics' ratings $(r(33) = .33, p = .05)$, suggesting both groups experienced similar, but not identical reading difficulties. Dyslexics appeared to spend longer time reading webpages than average readers $(83s$ vs $72s, t = 1.53, p = .06)$. However, inspection of reading-time histograms revealed a number of outliers: the participants would get distracted and come back to the task after a pause. In such case, a comparison of means was still possible (the t-test above), but the reading-time data could not be used in correlational analyses and we focused on the readability ratings instead.

We correlated the readability scores with the expert scores of compliance with guidelines. For 15 guidelines, the compliance scores did not satisfy the normality requirement for the Pearson product moment correlation. We used Spearman correlation coefficients instead. The correlation between compliance scores and readability scores was significant for five guidelines (Table 4), suggesting that following these five guidelines could improve readability.

Guideline	Dysl	Avg R	Corr
- Use short, simple sentences in a direct style.	-.39*	-.35*	P
- Avoid complex language and jargon.	-.50**	-.35*	P
-Consider using short paragraphs.	-.15	-.35*	P
- Put the main point of sentence or paragraph into the beginning of the sentence or paragraph.	-.46**	-.30	P
- Use section headings to organize the content.	-.36*	.20	S

** p < .01; * p < .05

Table 4. Correlations between readability scores and guideline compliance scores. The last column shows the correlation method (S – Spearman's; P – Pearson's).

The readability ratings of both dyslexics and non-dyslexics did not correlate with any WCAG-based automatic measures, neither with the number of detected accessibility errors nor with the number of errors per unit of webpage size. We used Spearman's rank correlation since the number of errors was not normally distributed.

The automatic measures of the guidelines (Table 3) performed better than the WCAG-based measures. Table 5 shows correlations between the measure scores and readability scores. The scores for two measures – text style and text alignment – were not normally distributed, which forced us to use Spearman's correlation coefficients instead of Pearson's coefficients. A review of Font Type scores revealed that most webpages used almost exclusively sans-serif fonts, ruling out a correlational analysis. Instead, the

numerical scores were converted in a factor: webpages with >30% of texts in Serif fonts were marked 'serif'; others 'sans-serif'. A t-test revealed a difference between the two webpage groups for dyslexics $(t = -1.65, p = .07)$, but not for average readers $(t = -.36, p = .37)$, which implied dyslexics struggled to read texts in serif fonts.

Measure	Dyslexics	Avg Readers	Corr
Amount of text	.18	.35*	P
Headers	-.46**	-.13	P
Text style: bold	-.01	.08	S
Text style: italic	.37*	-.03	S
Text style: underline	.38*	-.09	S
Font size	-.35*	-.11	P
Font type	--	--	--
Line spacing	.06	-.11	P
Width of text column	.34*	.14	P
Text alignment	.16	-.12	S
Luminance contrast	.10	.33*	P
Color difference	-.07	-.39*	P

Table 5. Correlations between readability ratings and automatic measures. The last column shows the correlation method (S – Spearman's; P – Pearson's).

Table 6 combines the guidelines that received empirical support from Table 4 and Table 5, except for *Luminance Contrast* (Table 5). The measure correlated positively with readability, implying higher contrast would increase reading difficulty, which contradicted the underlying guideline "*Ensure high luminance contrast between text and background, with the luminance of one 7 times the luminance of the other*". The contradiction could stem from people disliking black-on-white text, as said in "*Use an off-white color for your background, like light gray or tan; use dark gray for text instead of pure black*".

DISCUSSION

The research reported in this paper started from a large set of readability guidelines that could not be readily applied in Web design: they were too many and some of them were poorly formulated, too obscure, not validated, contradicted each other, or focused on text-to-speech software instead of reading behavior per se. We iteratively filtered out irrelevant guidelines and improved the others until the guideline set contained 61 actionable guidelines (Table 2).

We further reduced this 61-guideline set down to 12 core guidelines (Table 6) that our study indicated as effective. Interaction designers could rely on the validated 12 guidelines to quickly evaluate and iterate designs: 12 guidelines are relatively few and easy to follow, and at the same time, should suffice for a substantial improvement in webpage readability. The guidelines also bear implications for teaching. They give concise, actionable, and validated advice, which might ease learning design and accessibility, and help spread awareness of good design practices.

	Guideline	Dyslexics	Avg Readers
	- Use short, simple sentences in a direct style.	x	x
	- Avoid complex language and jargon.	x	x
E	*- Consider using short paragraphs.*		x
	- Put the main point of sentence or paragraph into the beginning of the sentence or paragraph.	x	
B	*- Use section headings to organize the content.*	x	
	- Limit the amount of content on a page to avoid scrolling.		x
	- Avoid using italics in the main body of the text.	x	
	- Avoid underlining large blocks of text as it makes reading harder.	x	
A	*- Use a minimum of text size 12pt or 14pt.*	x	
	- Avoid formatting texts in large-width columns.	x	
	- Use an off-white color for your background, like light gray or tan; use dark gray for text instead of pure black.		x
	- Use a plain, evenly spaced sans serif font such as Arial and Comic Sans.	x	

Table 6. The guidelines that received empirical support, based on experts' scores (E) or on automatic measures (A) or on both (B).

The research indicated that the experts could not evaluate with sufficient accuracy the low-level detail of webpages: none of the guidelines that the expert-based evaluation found helpful (Table 6, E) addressed such low-level visual detail. On the other hand, the algorithms could evaluate the low-level visual detail with sufficient accuracy (Table 6, A). Such observation may suggest the experts – all of whom were non-dyslexics – viewed the 'acceptable' levels for the low-level design detail differently from dyslexics and even from average readers. The experts may have failed to empathize with the user. This reinforces the participatory-design idea of actively involving the user in design process, but also suggests that the algorithms may act as a stand-in for the users, particularly for the user with accessibility disabilities.

The research also indicates that the WCAG 2.0 hardly addresses readability: from the core 12 guidelines, WCAG 2.0 seem to only include two, "*Use section headings to organize the content*" and "*Avoid formatting texts in large-width columns*". The next iteration of WCAG should include additional guidelines that specifically address reading. The automatic measures we developed in this paper (Table 3) could help to improve the next iterations of WCAG: they detail some of the guidelines and make them non-generic, which WCAG favors. These measures could also be used to develop an automatic validation approach

The study suggested that the readability issues of dyslexics differed from the readability issues of average readers: out of 12 core guidelines, only two applied to both dyslexic and average readers (Table 6). However, no guideline that helped dyslexics harmed average readers and vice-versa. All 12 guidelines can be applied to improve readability for all user groups and designers do not have to prioritize designing for one group at the expense of the other.

The study may also benefit fundamental dyslexia research, as some of 12 guidelines that we propose had not had empirical support. For example, three guidelines – avoiding underlining, using section headings, and putting the main

point as the first sentence – were neither tested nor actively discussed in the psychological research on dyslexia. Two more guidelines – avoiding wide text columns and avoiding italics – were tested [20] but without a control group, which would not let one conclude if the guidelines affected both dyslexics and average readers, or only dyslexics. Contrasting what phenomena affected dyslexics and what phenomena affected regular readers (Table 6) may help future research to better understand dyslexia and its outcomes.

The study contained several limitations that future research should address. First, future research will use more ecologically valid stimuli (e.g., websites instead of screenshots) and more realistic tasks (e.g., search instead of reading). Such stimuli and tasks may reveal the reading shortcuts that users might take (e.g., using a search function may let the user avoid reading an entire page). Second, future research will sample more webpages. Studying more webpages might reveal additional guidelines that should be included in the core set. Finally, future research will validate our results and conclusions, relying also on non-subjective data such as reading speed and number of errors or eye-tracking data.

CONCLUSION

This paper contributes two sets of Web readability guidelines. The first set (Table 2) includes 61 guidelines that have passed all three rounds of our guideline review. These 61 guidelines can potentially improve readability, but may be too many to be practical in many design settings. The second set (Table 6) includes 12 core guidelines that the study has highlighted as significantly effective for typical webpages. However, the 12 guidelines address design aspects present in many interfaces, and we expect them to apply to other-than-webpages interactive systems.

REFERENCES

1. A. Ailesgain, O'Connor S., D. O'donoghue, A. Quinn, and G. Tinney. (2011) Make it Easy: A guide to preparing Easy to Read information. [Online]. http://www.walk.ie/userfiles/file/Make%20It%20Easy%20-%20A%20guide%20to%20preparing%20Easy%20to%20Read%20Information.pdf

2. A. Al-Wabil, P. Zaphiris, and S. Wilson, "Web navigation for individuals with dyslexia: an exploratory study," in *International Conference on Universal Access in Human-Computer Interaction*, 2007, pp. 593-602.

3. American Psychiatric Association, *Diagnostic and statistical manual of mental disorders*, 5th ed. Washington, D.C: American Psychiatric Association, 2013.

4. F. A. Aziz and H. Husni, "Interaction Design for Dyslexic Children Reading Application: A Guideline ," in *Knowledge Management International Conference (KMICe)*, 2012, pp. 682-686.

5. British Dyslexia Association. (2012) Dyslexia style guide. [Online]. http://www.bdadyslexia.org.uk/common/ckeditor/filemanager/userfiles/About_Us/policies/Dyslexia_Style_Guide.pdf

6. C. J. Chen, M. W. Y. Keong, C. S. Teh, and K. M. Chuah, "Web text reading: what satisfy both dyslexic and normal learners?," *Journal of Computers in Education*, vol. 3, no. 1, pp. 47-58, 2016.

7. M. G. Friedman and D. N. Bryen, "Web accessibility design recommendations for people with cognitive disabilities," *Technology and Disability*, vol. 19, no. 4, pp. 205-212, 2007.

8. S. He, P. Cavanagh, and J. Intriligator, "Attentional resolution and the locus of visual awareness," *Nature*, vol. 383, pp. 334-337, 1996.

9. J. E. Jackson, "Towards Universally Accessible Typography: A Review of Research on Dyslexia," in *29th Annual International Technology and Persons with Disabilities Conference*, 2014.

10. T. B. Kirkwood, J. Bond, C. May, I. McKeith, and M. M. Teh, "Foresight mental capital and wellbeing project.," The Government Office for Science, 2008.

11. S. Kurniawan and P. Zaphiris, "Research-Derived Web Design Guidelines for Older People," in *7th ACM Conference on Computers and Assessibility*, 2005, pp. 129-135.

12. M. Martelli, G. Di Filippo, D. Spinelli, and P. Zoccolotti, "Crowding, reading, and developmental dyslexia," *Journal of Vision*, vol. 9, no. 4, pp. 1-14, 2009.

13. J. Nielsen. (2005) Lower-Literacy Users: Writing for a Broad Consumer Audience. [Online]. https://www.nngroup.com/articles/writing-for-lower-literacy-users/

14. E. Paulesu et al., "Dyslexia: cultural diversity and biological unity," *Sience*, vol. 259, no. 5511, pp. 2165-2167, 2001.

15. M. Perea and V. Panadero, "Does Viotin Activate Violin More Than Viocin?," *Experimental Psychology*, vol. 61, no. 1, pp. 23-29, 2014.

16. H. Petrie and O. Kheir, "The relationship between accessibility and usability of websites," in *SIGCHI conference on Human factors in computing systems* , 2007, pp. 397-406.

17. E. Pitler and A. Nenkova, "Revisiting readability: A unified framework for predicting text quality," in *Conference on Empirical Methods in Natural Language Processing*, 2008, pp. 186-195.

18. P. Rainger. (2003) A Dyslexic Perspective on e-Content Accessibility. [Online]. https://users.dimi.uniud.it/~giorgio.brajnik/dida/psw/galleria/05/vallega/relaz-1/03-04-01_dyslexic.pdf

19. L. Rello and R. Baeza-Yates, "Good fonts for dyslexia," in 15th International ACM SIGACCESS Conference on Computers and Accessibility, 2013.

20. L. Rello, G. Kanvinde, and R. Baeza-Yates, "Layout guidelines for web text and a web service to improve accessibility for dyslexics," in *International cross-disciplinary conference on web accessibility*, 2012.

21. L. Rello, M. Pielot, and M. C. Marcos, "Make it big! The effect of font size and line spacing on online readability," in *CHI'16*, 2016, pp. 3637-3648.

22. L. Rello, M. Pielot, and M.C. Marcos, "Size Matters (Spacing Not): 18 Poionts for a Dyslexic-friendly Wikipedia," in *10th International Cross-Disciplinary Conference on Web Accessbility*, 2013.

23. G. Rugg and P. McGeorge, "The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts," *Expert Systems*, vol. 22, no. 3, pp. 94-107, 2005.

24. V. F. Santana, R. de Oliveira, L. D. A. Almeida, and M. C. C. Baranauskas, "Web accessibility and people with dyslexia: a survey on techniques and guidelines," in *International Cross-Disciplinary Conference on Web Accessibility*, 2012.

25. V. F. Santana, R. Oliveira, L. D. A. Almeida, and M. Ito, "Firefixia: An accessibility web browser customization toolbar for people with dyslexia," in *10th International Cross-Disciplinary Conference on Web Accessibility*, 2013.

26. M. H. Schneps et al., "Shorter lines facilitate reading in those who struggle," *PloS one*, vol. 8, no. 8, 2013.

27. P. H. K. Seymour, M. Aro, and J. M. Erskine, "Foundation literacy acquisition in European orthographies," *British Journal of Psychology*, vol. 94, pp. 143-174, 2003.

28. D. Spinelli, M. De Luca, A. Judica, and P. Zoccolotti, "Crowding effects on word identification in developmental dyslexia," *Cortex*, vol. 38, no. 2, pp. 179-200, 2002.

29. M. F. Theofanos and J. Redish, "Guidelines for accessible and usable web sites: Observing users who work with screen readers," *Interactions*, vol. 10, no. 6, pp. 38-51, 2003.

30. A. Tseng. (2012) uxmovement.com. [Online]. http://uxmovement.com/content/ 6-surprising-bad-practices-that-hurt-dyslexic-users/

31. T. R. Vidyasagar and K. Pammer, "Dyslexia: a deficit in visuo-spatial attention, not in phonological processing," *Trends in Cognitive Sciences*, vol. 14, no. 2, pp. 57-63, 2010.

32. D. Whitney and D. M. Levi, "Visual crowding: A fundamental limit on conscious perception and object recognition," *Trends in cognitive sciences*, vol. 15, no. 4, pp. 160-168, 2011.

33. World Wide Web Consortium. (2008) w3.org. [Online]. https://www.w3.org/TR/WCAG20/

34. Y. Yeshurun and E. Rashal, "Precueing attention to the target location diminishes crowding and reduces the critical distance," *Journal of Vision*, vol. 10, no. 10, pp. 1-16, 2010.

35. C. H. Yu and R. C. Miller, "Enhancing web page readability for non-native readers," in *the sIGCHI conference on human factors in computing systems*, 2010, pp. 2523-2532.

36. V. Zarach and JISC CETIS. (2002) Ten Guidelines for Improving Accessibility for People with Dyslexia. [Online]. http://wiki.cetis.ac.uk/Ten_Guidelines_for_Improving_Accessibility_for_People_with_Dyslexia

37. M. Zorzi et al., "Extra-large letter spacing improves reading in dyslexia," *Proceedings of the National Academy of Sciences*, vol. 109, no. 28, pp. 11455-11459, 2012.

38. Braun, C. C., Mine, P. B., & Silver, N. C. (1995). The influence of color on warning label perceptions. *International Journal of Industrial Ergonomics*, *15*(3), 179-187.

39. K. Höök and J. Löwgren. Strong concepts: Intermediate-level knowledge in interaction design research. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *19*(3), 23, 2012.

40. E. Pitler and A. Nenkova. "Revisiting readability: A unified framework for predicting text quality." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 186-195). Association for Computational Linguistics, 2008.

41. N. C. Silver and C. C. Braun. "Perceived readability of warning labels with varied font sizes and styles." *Safety Science*, *16*(5-6), 615-625, 1993.