

U-index: An Eye-Tracking-Tested Checklist on Webpage Aesthetics for University Web Spaces in Russia and the USA

Svetlana S. Bodrunova^(✉) and Alexander V. Yakunin

St. Petersburg State University,
7-9 Universitetskaya nab., St. Petersburg 199004, Russia
{s.bodrunova, a.yakunin}@spbu.ru

Abstract. *Background.* Understanding the relations between user perception and aesthetics is crucial for web design. But it is frequent in today's graphic and media design that rules, established by practitioners even before the advent of Internet and still untested empirically, are taught at design schools and widely used for online interface design. So far, there is no well-established linkage between the in-class recommendations and our empirical knowledge on usability, for which design plays a role just as crucial as web projecting. Will webpages that are better from the designers' viewpoint perform better in terms of usability? And can one have a list of recommendations tested empirically?

This is especially important for large-scale organizational web spaces where design plays a huge role in brand recognition and visual unity. Large web spaces need complex ergonomic assessment both on the level of selected nodes and on that of architecture/navigation. Of many large web spaces, university portals suit best for elaboration and pre-testing of such a methodology, as they serve various publics, contain sub-domains, and often face criticism for their user-unfriendly design and messy structure.

Objectives. We aim at creating a two-level usability expert test for a large web space that would be based on design recommendations tested empirically, thus eliminating the necessity of tech-based assessment of newcoming products. In this paper, we elaborate the node-level methodology. For this, basing on leading design literature, we create a page usability index (U-index) for 'good' design that provides quantitative measurement for traditional design decisions on the micro- and macro-level of a web page. Then, we test by eye tracking whether 'better' design (corresponding to higher U-index values) favors a particular pattern of content consumption – not 'random search' but more efficient 'reading'.

Research design. To check whether web design measured qualitatively correlates with perception of web pages as tested by eye tracking, we first define target nodes by collecting the hyperlink structure and constructing web graphs for three web spaces of the biggest universities in the USA and Russia (Harvard University, Moscow State University, St. Petersburg State University). For this, we combine web crawling and web analytics. Second, we construct the U-index with the maximum value of 22. Third, to assess user perception of the target web pages, we create a series of tasks on information search and measure three test parameters (number of eye fixations, duration of fixations, and saccade length) and their derivatives, as well as heat maps. To avoid bias in quantitative

measurement, we use two eye trackers (one head-fixed, one stationary) to test the results in parallel. Fourth, for finding correlations between U-index and eye-tracking results, descriptive statistics (Spearman's rho and Cramer's V) is used.

Results. First of all, our results suggest that various types of eye tracking hardware produce very different test results; this implies that eye tracking research always needs pre-testing. Second, we see that heat maps may be very suitable in express assessment of the web design quality, which speaks in favor of preserving some eye tracking tests in the final methodology. Third, we see substantial difference between Russian universities and Harvard: the latter, indeed, shows that features of web design correlate with eye tracking experience of the assessors, while for the Russian university websites, even after their repeated attempts of redesign, it remains unclear whether web design contributes to better user experience. For Harvard, the web pages with a higher usability index tend to facilitate 'reading' instead of 'search'. Fourth, micro-level elements of the layout seem to contribute more to the general index and, thus, may deserve bigger attention of web designers.

Keywords: Web usability · Web design · Aesthetics · Eye tracking · U-index

1 Literature Review

Algebra and harmony: linking aesthetic parameters of web design to user experience. Inter-relation between web usability and web aesthetics is believed to be one of the major research areas within web usability studies (Hassenzahl and Monk 2010; Bargas-Avila and Hornbæk 2011). In 1990s – early 2000s, the first 'aesthetics vs. interface usability' studies by Kurosu and Kashimura (1995) and Tractinsky et al. (2000) showed that visual appearance casts independent impact upon user perception of their interaction with interfaces. Later, many studies showed the influence of aesthetics on perception of usability (Ben-Bassat et al. 2006; Thüring and Mahlke 2007) and overall user impression (Schenkman and Jönsson 2000); the latter work even stated that the best predictor for the overall user judgment on a website was its aesthetic appeal. But till today, there is no clear answer whether and, more important, in what way aesthetics on the whole or its particular aspects are linked to web usability. As Tuch et al. (2012) have noted, several later studies have proved the linkage itself, while others showed there was no direct relation between the two.

There are, to our viewpoint, several reasons for the absence of clear answer to this 'beautiful and/or usable' dilemma – despite the fact that there is growing evidence of the positive answer to it (Sonderegger and Sauer 2010). At least partly, it lies in the 'lack of experimental studies manipulating aesthetics and usability as independent variables' (Tuch et al. 2012, p. 2). Also, most works that explore the linkage and may state causal relations between aesthetics and usability, use the data that are correlational by nature (Tractinsky et al. 1997; van Schaik and Ling 2003, 2008; Hassenzahl 2004; De Angeli et al. 2006), and 'the causality is solely a matter of theoretical reasoning and cannot be tested by existing data' (Tuch et al. 2012, p. 2). While we do see a problem with causality/correlations, we would argue that the first problem – the lack of

experimental studies – seems to be a bigger one. In our viewpoint, it partly lies in the fact that, today, web design comprises the areas of design activity as different as prototyping, layouting, graphic design, and web architecture&navigation. Partly due to this, it seems hard to establish a clear set of measurable variables that could be tested via objective means like eye tracking.

The majority of works that claim testing the web design actually test architectural and navigation problems. This comes from early academic works on web design (see Goldberg et al. 2002 as one of the earlier examples setting this trend), but in many today's web agencies prototyping/layouting and graphic design for web are considered different professions; working with page structural elements lies somewhat in between, as the set of necessary elements is defined by account managers and prototypists, while their visual appearance and relative visual salience is the designers' area of responsibility. But this duality in the understanding of what is actually analyzed by eye tracking tests is virtually never addressed in academic works. Thus, a line of studies have focused on menu layouting (McCarthy et al. 2004; Leuthold et al. 2011) as a 'design' element. Similarly, some studies have looked at the page elements that are most attractive for specific audiences, calling them 'design elements', which are actually page structure ones. Thus, Djasasbi et al. (2010) have shown that 'a main large image, images of celebrities, little text, and a search feature' are the design elements that attract Generation Y audience. But among the elements named, there are no graphic design ones as such. Many works assess layouting features (Buscher et al. 2009).

A sub-area that comes closer to linking graphic design and web usability testing is heuristic evaluation (Nielsen 1994; De Kock et al. 2009). The works in this research zone provide multiple criteria for heuristic evaluation (see the Multiple Heuristic Evaluation Table excerpts, De Kock et al. 2009, p. 131). But as this method is based on expert evaluation, the way the assessment criteria are formulated ('all the text in black font is easy to read', 'the graphics convey information clearly') cannot be applied for objective measurement of design features.

Also, there's a range of works that test the so-called first user impression; in the 2000s, it was shown that the 'good' design (as measured by users) raises the level of user satisfaction (but not the efficacy of information search; Phillips and Chaparro 2009). The authors even claimed that visual appearance had a long-lasting impact upon user satisfaction, since design perceived as better helped maintain the positive impression even after user interactions with manipulated pages. Another study (Lindgaard and Dudek 2002) also found that, even if user satisfaction dropped significantly after running flawed web pages, the aesthetic perception remained high. Reinecke et al. (2013) inter-changed the independent and dependent variables and tried to predict users' first impressions by correlating them to perceived visual complexity and colorfulness; a similar effect on 'what is usable is beautiful' was discovered by Ilmberger et al. (2008). But these studies are united by seeing aesthetics as just 'high/low', 'good/poor' (e.g., Moshagen et al. 2009). Even the famous study by Lavie and Tractinsky (2004), despite complicated methodology, results in measures like 'clean', 'symmetrical', 'pleasant', or simply 'aesthetic' design. Studies in related areas also used the 'attractive/non-attractive' division, rather than measurable parameters (see, e.g., Sonderegger and Sauer 2010; Quinn and Tran 2010).

Some works, though, provide a closer look at some single design elements. Thus, in their famous study ‘Eyetrack III’ of early 2000s, the Poynter Institute have hinted that smaller case and shorter headlines would lead to higher usability results (Outing and Rual s.d.). Lindgaard (2007), as many other colleagues afterwards, focused on color and color combinations, while Cyr et al. (2010) have stated cross-cultural differences in the impact of color on user experience. Bernard, Chaparro and Thomasson, as early as in 2000, have stated that whitespaces amounts play a role for subjective user satisfaction, rather than for task performance; later works (Coursaris and Kripintris 2012) further proved the mid-level of whitespaces as a significant factor in raising user experience data.

More systemic works that try to catch a variety of aesthetic features of web design are, i.e., found in two other literature streams. One of the research areas worth looking at is the one on visual complexity of websites and various types of images. Thus, we take into account the work by Pieters et al. (2010) where they state and describe the ‘design complexity’ (as distinguished from ‘feature complexity’), as well as show that design complexity raises, not lowers, the efficacy of ad perception. The criteria of design complexity provided by the authors are, again, non-measurable; but the authors make an attempt to measure it in terms of yes/no, and we will partly follow this logic. But perhaps the stream of literature and experiments that comes closest to objectivization of web aesthetics is the one that derives from the works by Ngo and colleagues (Ngo 2001; Ngo and Byrne 2001; Ngo et al. 2000, 2003 and others). Thus, based on them, Purchase et al. (2011) have elaborated 14 aesthetic parameters like balance, equilibrium, symmetry etc., all measured (0:1), and showed their relevance for user experience. This work is a rare attempt to bring objective measurement into the highly subjective area of aesthetics.

But what these works still lack is the relation between specific graphic design and layouting recommendations, on one hand, and user experience, on the other. Till today, qualitative assessment of web pages remains largely detached from the literature for graphic designers and web designers where practical advices in color, spacing, lineage and other artistic aspects were discussed by design gurus. This paper aims at finding correlations between qualitative assessment of design of web pages and eye-tracking results for the same web pages, thus (possibly) linking the existing tradition of graphic design to today’s understanding of web usability and its metrics. In other words, we would like to test whether recommendations provided in today’s design manuals are, indeed, relevant and may be combined in a checklist of sustainable and testable recommendations.

University websites as the testing ground: 5-step usability assessment for large organizational web spaces. Among various types of websites, large web spaces (e.g. web portals of large organizations) remain under-researched in terms of user satisfaction in both ergonomics and web architecture and navigation. Today, analysis of web efficacy has reached the level when it is possible to analyze not only individual web pages but large web segments that include hundreds of thousands of pages; in most cases, they can be reconstructed as networks via web crawling (Blekanov et al. 2014). They represent web clusters where the same design pattern needs to be replicated but also necessarily changes for sub-domains and within the page hierarchy.

Thus, large web spaces, especially web portals of large organizations, may represent suitable objects for elaboration and testing of comparative methodology of assessment of efficacy of web design. In this research zone, web analytics is intertwined with design, engineering psychology, and micro-ergonomics. Thus, for such research objects, web metrics and usability metrics should be viewed as two inter-dependent and interweaved sets where dependencies are still to be checked and tested.

Among large web spaces, university web spaces represent a special cluster and suit well for our analysis, as they serve very different publics, are multi-task, contain sub-domains and often evoke criticism for their messy structure and user-unfriendly design. Moreover, they are available to the researchers full-time, which made university websites a popular research object. University websites have become a usual object for both single-country (Zaphiris and Ellis 2001; Hasan 2012) and cross-cultural comparative web usability tests. Our novelty here is that we treat the university website as a web space where the basic design elements are responsible for brand recognition despite the differences between schools and colleges inside a university, as well as presence of many additional pages within university web clusters. For this paper, we still focus on the core websites of the universities (the main-domain sites) but aim at expanding the research to the ‘web space’ level.

We have selected the web spaces of Moscow and St. Petersburg State Universities as the two largest universities in Russia, and also Harvard University website known for its minimalistic design; we expect the latter to work in a way as a benchmark for our assessment of the two Russian universities.

2 Research Design and Methodology

For large web spaces, we propose a complex usability test based on three steps. Step 1: selection of key nodes for analysis (by web crawling and web analytics), Step 2: ergonomic (page-level, or node-level) tests, Step 3: architecture and navigation tests (for the detailed account, see Bodrunova et al. 2016). In this paper, we focus on Step 2 and use Step 1 for page sampling. Also, in order to pre-test the suggested methodology, we will introduce a comparative component into the study.

A special role in any web space belongs to key architectural and network nodes; their usability and navigation through them must be looked at with special attention. We will focus on testing the key nodes comparable in terms of their position in graphs and the main website menus (that is, the nodes that were meant to become key pages and are, indeed, performing these roles).

For Step 1 (selection of web pages), a web crawler with specialized modules was tested (Blekanov et al. 2012) and adapted. As a result of web crawling for the three universities, three web graphs were reconstructed. Based on them, we have singled out comparable pages. We have included into pre-testing high-rank web pages only, as based on their network centrality data: a page had to belong to top pages by at least one of SNA centrality metrics for all three universities, to be comparable in its overall structure and aims to its counterparts and, third, to play an important role in the structure of the university website.

The following pages were chosen (15 pages altogether): homepage; university news; university structure (+ contacts of the faculties/institutes/personnel); scientific life (coverage of main scientific events); university life (announcements and short news-like coverage).

For Step 2a (qualitative assessment). Despite their importance, elements of graphic design traditionally considered responsible for readability and user friendliness of a media interface (and used, e.g., in newspaper layouting), were not tested well enough, neither individually nor as a complex. These elements are parts of the so-called composite-graphic model (further referred here as CGM), or, in a simpler and less precise way, a layout (which is the result of application of the CGM to a particular page). To overcome the term mess existing in today's literature and described above, we will use CGM as a referenced-to whole that comprises page-level and element level components. We state that there is no agreed methodology of CGM usability testing.

As stated in Sect. 1, most researchers focus on just one or several aspects of visual organization of a page and formulate dependencies that relate user experience to particular elements of CGM. But we take into account the integrity of a web project; this implies a certain hierarchy in designers' decision-making. Thus, based on works by Velichkovsky (2010), we have divided CGM's visual organization into two levels. The macro-level comprises composition, color, zonation, and page-level spacing and deals with heterogeneity, content combination, and visual saliency of layout elements; micro-level comprises individual-block parameters, typography, inter-line spacing, and syntagmas and deals with readability and cognition speed.

For single page assessment, we have elaborated a qualitative index of usability for a web page (GCM usability index, or U-index) based on a wide range of literature on traditional newspaper design, digital news design, web design, and perception theory. We focus not only on the works of graphic design gurus as Arnheim or Berlyne but also on over a dozen today's popular design and web design manuals.

In the U-index, we have also combined the findings of those who analyzed the layouting features and the scarce findings on the web graphic design. On the macro-level, we have taken into account notion of website visual complexity and cognitive load (Pieters et al. 2010; Wang et al. 2014), as website complexity seems to play a role exactly for the mode of webpage consumption: high-complexity websites, counter-intuitively, tend to facilitate the 'reading' mode (see below; Wang et al. 2014). In the test tasks, both levels were taken into consideration.

The index includes the following categories:

- macro-level: overall type of layout; layout module structure; vertical spacing; page zonation; creolization of the layout;
- micro-level:
 - syntagma: line length, line length in title block, leading (inter-lineage spacing);
 - typography: contour contrast, tone&color contrast with background, font adaptivity, x-height, font&line length combination.

Each of the 13 chosen parameters were given values (0; 1) or (0; 1; 2). The overall maximum index of a page equals to 22.

Our goal is to test the U-index on the whole, as well as on the macro-and micro-levels. In future, individual parameters of the U-index are to be tested; but we first wish to prove that there is, in general, a link between user experience and the index.

For Step 2b (quantitative test), we have elaborated the information search tasks for each page based on the existing practices of usability testing (Broder 2002; Rose and Levinson 2004), as well as on the pre-assumption that there are two modes of a user's interaction with a page: that of *search* and that of *reading*, as suggested by Velichkovskiy (2010). The mode of 'reading' is desirable, while the mode of 'search' is not, as in the existing studies 'reading' implies focused studying of target content, while 'search' implies random looking for it on a page. The tasks we designed were oriented to finding a piece of target content, not to in-depth understanding or long-term memory on it, as such tasks let assess how quickly 'search' transforms into 'reading' and whether 'reading' dominates. Thus, five tasks were elaborated, each one adapted for three different pages; we made sure they were comparable in each case. So far, the task complexity as an intervening factor was not tested within our pre-test.

The testing methods we use are heat maps and metrics of eye movement. Among the latter, eye fixations are assessed, most often by three metrics (Salvucci and Goldberg 2000; Poole et al. 2004): the number of fixations, their duration, and saccade length (the distance between two fixation dots on the monitor). We explore the 'search'/'reading' modes by these metrics as well as their derivatives.

Two new derivative metrics that we suggest are calculated as mean deviations of the main eye movement metrics. Thus, while mean fixation length provides hints on overall mean readability of the page, mean fixation length deviation tells whether the layout is the same throughout the page, or some parts of the content are consumed faster than others. Similarly, mean saccade length deviation hints to the dynamics of 'search'/'reading', thus enriching our knowledge on the overall content consumption pattern described by the average saccade length. Lower deviations would tell of the 'reading' pattern, while higher deviations would be a sign of 'search'.

Then, we have introduced several more comparative elements into our research design and sampling.

First, as several recent works have stated that only fixations over 300 ms count, we created two datasets for each university – namely, the one with all the fixations of eye motion and the one with the fixations of 300+ ms. Thus, we will analyze the eye tracking results comparing the samples of all fixations and of fixations of 300+ ms.

Second, we will compare the results of the two eye tracking methods: the quantitative measures (number of fixations, fixation duration, saccade length, and the derivative metrics) vs. heat map assessment. We further elaborated the latter quantitatively and included five metrics in our study:

- overall number of red spots on the screen (in N -> 0; 1; 2);
- number of red spots close to the target element (in N -> 0; 1; 2);
- size of the maximal red spot closest to the target element (in mm -> 0; 1; 2);
- intensity of the biggest red spot closest to the target element (qualitatively -> 0; 1);
- diameter of the maximal red spot closest to the target element (in mm -> 0; 1; 2).

Third, to be able to recommend the U-index to professional designers, we wanted to ensure that the results of assessment are not hardware-dependent and that any eye tracker would produce similar results. Two eye trackers (one stationary and one ‘unobtrusive’ with head fixation) were used. Two groups of four assessors each performed the same search tasks on one of the two eye trackers.

Thus, as heat maps were tested on the first eye tracker only, for each of the 15 pages chosen, 13 qualitative and 10 to 15 quantitative/mixed-method variables were assessed. Of the 13 variables in U-index, 3 figures were formed (for micro-level, macro-level, and their combination). Thus, the pre-test sample includes 40 entries with a single web page as the unit of analysis; 20 entries were measured by 18 variables each, and other 20 by 13 variables (excluding heat map analysis). All data on eye movement were assessed twice – within all-fixation and 300+ ms samples. For the resulting variables and research design overview, see Table 1.

Table 1. Testing the U-index: the research design and variables

U-index, general (Uigen)	U-index, macro-level (Uimac)	U-index, micro-level (Uimic)
-----------------------------	---------------------------------	---------------------------------

vs.

Variable category	Variable sub-category	Fixations		Variable category	All fix.
		All	300+		
Number of fixations	Number of fixations (Nfixall)	x	x	Overall N of red spots (Nrspots)	x
Fixation duration	Average fixation duration (Dfixall)	x	x	N of red spots near target (Ntar)	x
	Mean duration deviation (Dfixmall)	x	x	Max red spot size (Maxsize)	x
Saccade length	Average saccade length (Savall)	x	x	Max red spot intensity (Maxint)	x
	Mean length deviation (Sdevall)	x	x	Max red spot diameter (MaxD)	x
Eye tracker 2					
Variable category	Variable sub-category	Fixations		Variable category	All fix.
		All	300+		
Number of fixations	Number of fixations (Nfix300)	x	x		
Fixation duration	Average fixation duration (Dfix300)	x	x		
	Mean duration deviation (Dfixm300)	x	x		
Saccade length	Average saccade length (Sav300)	x	x		
	Mean length deviation (Sdev300)	x	x		

After the eye tracking test, we have looked for correlations between independent (U-index) and dependent (eye tracking) variables applying two different statistical metrics – Spearman’s correlation metric and Cramer’s V cross-tabulation metric.

3 The Research Hypotheses

H1. We consider eye trackers to be equivalent in their capacity of measuring user-interface interaction efficacy. Thus, two assessor groups will provide similar results.

H2. We consider web design to cast impact upon user interaction experience. Thus, U-index (measured on micro-level, macro-level, and on the whole) will correlate with user experience metrics.

H3. In our opinion, efficient design should facilitate the ‘reading’ mode. Thus, we hypothesize that, on the pages with better CGM (that is, with higher U-index), subjective user experience will be more like ‘reading’, not like ‘search’. That is, it will have:

- lower number of fixations;
- lower mean fixation duration;
- lower mean fixation duration deviation;
- lower mean saccade length;
- lower mean saccade length deviation;
- bigger and more intense ‘heat’ grouped around the target elements.

H4. Despite the existing literature, we do not expect the results for all-fixation sample to differ from the 300+ ms one.

H5. Due to its minimalistic design, Harvard will perform better than the Russian universities in all the aspects.

4 Conduct of Pre-test

For the pre-test, each assessor was asked to conduct 15 tasks (one task for each page), where the tasks for each page type were identical. Average session duration was between 30 and 40 min. As the tasks were similar but language-dependent, all the assessors were native Russian speakers with good command of English as well (EILTS 6 or higher). The eye tracking procedures took place in soundproof rooms. The same supervisor assisted at the procedures. The groups were homogeneous in terms of age (Master students) and were slightly familiar with all the three websites before the test.

5 Results

All in all, application of descriptive statistical metrics has returned the following results (only significant, marked bold, and slightly insignificant, marked italic, values are included). Please see Table 2 for Spearman correlation for groups 1 and 2, and Table 3 for Cramer’s V for groups 1 and 2.

H1 proves to be wrong. As evident from the Tables 2 and 3, we have discovered high differences in the results that we received from the two groups of assessors. This brings in new premises for future eye tracking research of web interfaces, as it is not

Table 2. Spearman correlations for groups 1 and 2

Group 1:

		Nfixall	Nfix300	Dfixall	Dfixmall	Dfix300	Dfixm300	Savall	Sdevall	Sav300	Sdev300	Nrspots	Niar	Maxsize	Maxint	MaxD
Harvard	Uigen											0,483*	0,722***			0,504*
	Uimac												0,628***			0,659***
	Uimic	-0,387 (<i>p</i> = ,092)						0,393 (<i>p</i> = ,086)					0,605**			
MSU	Uigen															
	Uimac															
	Uimic															
SPbU	Uigen												-0,411 (<i>p</i> = ,072)			
	Uimac															
	Uimic												-0,384 (<i>p</i> = ,095)			

Group 2:

		Nfixall	Nfix300	Dfixall	Dfixmall	Dfix300	Dfixm300	Savall	Sdevall	Sav300	Sdev300
Harvard	Uigen	-,747**	,644**	,761**		,547*	,776**	-,652**	-,667**		,736**
	Uimac	-,696**	,447*	,679**		,514*	,465*	-,471*	-,529*		
	Uimic	-,609**	,597**	,648**		,474*	,776**	-,699**	-,631**		,736**
MSU	Uigen										
	Uimac										
	Uimic										
SPbU	Uigen			,437 (<i>sig</i> .0,053)				-,467*			
	Uimac										
	Uimic	-,483*		,543*				-,549*			

only the quality of eye tracking data that are of concern (Holmqvist et al. 2012) but also the nature of the eye tracker itself that matters. Our suggestion for continuation of our own research is to use the ‘unobtrusive’ eye tracker – not because it has produced more substantial results but due to its unobtrusive nature.

H2, as we see, looks partly proven. Eye tracker 1 seems to reject it, except for the heat maps, and with the latter, only Harvard shows significant correlations for three of five variables. But eye tracker 2 shows that nearly all kinds of the suggested variables form significant correlations with the U-index for Harvard, and the traditional all-fixation-encompassing variables make it for SPbU; the results are also partly supported by Cramer’s V. This, first of all, provides the premises for future research, as the U-index seems to have relevance, at least in the case of one university (Harvard). Then, we need to know why MSU, SPbU and Harvard produced so different results; other factors, perhaps, need to be taken into consideration, as we definitely see that there is an overall cause of a dramatic difference between Harvard and MSU in terms of

Table 3. Cramer’s V cross-tabulation for groups 1 and 2

Group 1:

		Nfixall	Nfix300	Dfixall	Dfixmall	Dfix300	Dfixm300	Savall	Sdevall	Sav300	Sdev300	Nrspots	Ntar	Maxsize	Maxint	MaxD	
Harvard	Uigen																
	Uimac											.725*	0,569 (p=.086)				
	Uimic																
MSU	Uigen					,901 (p=.067)							.701**				
	Uimac												.683 (p=.069)				
	Uimic												.720**				
SPbU	Uigen																
	Uimac	.968 (p=.065)															
	Uimic	.968 (p=.060)															

Group 2:

		Nfixall	Nfix300	Dfixall	Dfixmall	Dfix300	Dfixm300	Savall	Sdevall	Sav300	Sdev300
Harvard	Uigen		.760***			.886**	.690***				.632**
	Uimac		.655**								
	Uimic		.849***			.968***	.690***			1.000***	.632**
MSU	Uigen										
	Uimac										
	Uimic										
SPbU	Uigen										
	Uimac		.901**								
	Uimic										

correlation between the U-index and assessors’ performance, be it the overall minimalist style of the Harvard web design or some other outer factor. But we also need to notice that language must, evidently be excluded as such a cause, as web design features seem to facilitate the assessors’ performance, but in English – contrary to expectations, as the assessors were Russian native speakers. So far, we can say that traditional metrics (that is, all-encompassing number of fixations, fixation duration, and saccade length) worked for two universities on eye tracker 2, and 300 + metrics did so only for Harvard, as well as deviation metrics introduced by us. At the same time, we cannot help noting that the mean deviation metrics also worked well for eye tracker 2, especially in case of 300+ ms metrics. This may mean that mean deviations may be

used to detect not only the direct efficacy of eye motion (e.g. timing of eye motion) but also more sophisticated patterns of content consumption.

H3 is also partly proven, and our results bear the logic of the ‘reading’ pattern, especially in case of Harvard. For this university:

for heat maps, *H3* is supported, as the U-index is higher if:
 more red spots are there on the screen;
 more red spots surround the target content;
 diameter of the major red spot near target is bigger.

For quantitative metrics, the situation is more nuanced than we hypothesized. Thus, indeed, higher U-index correlates with lower overall number of fixations, as well as lower saccade length and its mean deviation. That is, efficient pages are consumed with fewer ‘stops’ and smaller ‘jumps’ around the page. But at the same time, as soon as we take into account only ‘long stops’ (fixations of 300+ ms), the bigger number of fixations and saccade length, the better. Taken together with the fixation duration, these metrics form the ‘reading’ pattern, while big numbers of short fixations and short saccades form the ‘random search’ pattern.

H4, as seen from *H3*, is wrong; moreover, the case of Harvard shows that both overall and 300+ ms fixations metrics need to be taken into consideration in the usability tests, as they relate to different patterns of user-interface interaction, if taken together with other metrics.

H5 proves right on available data but definitely needs further research. Thus, on the whole and on micro-level, Harvard shows stronger correlations between qualitative assessment and eye tracking results. This may mean that Harvard needs smaller improvement of design to get the same user efficacy; also, design has a bigger chance to cast impact upon user-interface interaction within the university web space. Taking into consideration absence of clear picture for MSU and almost the same picture for SPbU, we conclude that design of the Russian university websites has smaller impact upon user experience and, thus, must be less efficient than that of Harvard.

6 Conclusion

So far, our results suggest that, at least in some cases, there is linkage between qualitative ‘designer’ understanding of efficient web design and user experience in content consumption. For Harvard, we have discovered that web pages more efficient from the designer viewpoint tend to have the ‘reading’ pattern (relatively small number of long fixations quite distanced from each other) and not the ‘search’ pattern (a lot of short fixations with short ‘jumps’ between them). This means that the U-index (after more testing and fine-grained research) may become a practical instrument for web designers and experts in practical web usability tests. Also, we have discovered the necessity to combine the variables to describe ‘search’ and ‘reading’ patterns in a more nuanced way.

But at the same time our results for the Russian universities show that web design might be irrelevant for assessors’ performance on search tasks; counter-intuitively, native Russian speakers demonstrated a more solid pattern of interaction with the American website, rather than with the Russian ones. Further research is needed to

define the factor that prevents web design from casting impact upon user experience in the latter cases.

Acknowledgements. This research has been supported in full by Russian Foundation for Basic Research, research grant 15-01-06105 (2015–2016) ‘Elaboration of webometric and ergonomic models and methods of analysis of efficacy of web presence of large organizations’.

References

- Bargas-Avila, J., Hornbæk, K.: Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, pp. 2689–2698. ACM (2011)
- Ben-Bassat, T., Meyer, J., Tractinsky, N.: Economic and subjective measures of the perceived value of aesthetics and usability. *ACM Trans. Comput. Hum. Interact.* **13**(2), 210–234 (2006)
- Bernard, M., Chaparro, B., Thomasson, R.: Finding information on the web: does the amount of whitespace really matter. *Usability News* **2**(1), 1 (2000)
- Blekanov, I.S., Sergeev, S.L., Martynenko, I.: Construction of subject-oriented web crawlers using a generalized kernel. *Sci. Tech. Bull. St. Petersburg State Polytech. Univ.* **5**(197), 9–15 (2012)
- Bodrunova, S.S., Yakunin, A.V., Smolin, A.A.: Comparing efficacy of web design of university websites: mixed methodology and first results for Russia and the USA. In: Proceedings of the International Conference on Electronic Governance and Open Society: Challenges in Eurasia, pp. 237–241. ACM (2016)
- Broder, A.: A taxonomy of web search. *SIGIR Forum* **36**(2), 3–10 (2002)
- Buscher, G., Cutrell, E., Morris, M.R.: What do you see when you’re surfing?: using eye tracking to predict salient regions of web pages. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 21–30. ACM (2009)
- Chaparro, B., Bernard, M.: Finding information on the web: does the amount of white space really matter? In: Proceedings of the Tenth Annual Usability Professionals’ Association Conference (2001). <http://psychology.wichita.edu/surl/usabilitynews/W/whitespace.htm>
- Coursaris, C.K., Kripintris, K.: Web aesthetics and usability: an empirical study of the effects of white space. *Int. J. e-Business Res. (IJEBR)* **8**(1), 35–53 (2012)
- Cyr, D., Head, M., Larios, H.: Colour appeal in website design within and across cultures: a multi-method evaluation. *Int. J. Hum Comput Stud.* **68**(1), 1–21 (2010)
- De Angeli, A., Sutcliffe, A., Hartmann, J.: Interaction, usability and aesthetics: what influences users’ preferences? In: Proceedings of the 6th ACM Conference on Designing Interactive Systems, pp. 271–280. ACM (2006)
- De Kock, E., Van Biljon, J., Pretorius, M.: Usability evaluation methods: mind the gaps. In: Proceedings of the 2009 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists, pp. 122–131. ACM (2009)
- Djamasbi, S., Siegel, M., Tullis, T.: Generation Y, web design, and eye tracking. *Int. J. Hum. Comput. Stud.* **68**(5), 307–323 (2010)
- Goldberg, J.H., Stimson, M.J., Lewenstein, M., Scott, N., Wichansky, A.M.: Eye tracking in web search tasks: design implications. In: Proceedings of the 2002 Symposium on Eye Tracking Research & Applications, pp. 51–58. ACM (2002)

- Hasan, L.: Evaluating the usability of nine Jordanian university websites. In: 2012 International Conference on Communications and Information Technology (ICCIT), pp. 91–96. IEEE (2012)
- Hassenzahl, M.: The interplay of beauty, goodness, and usability in interactive products. *Hum. Comput. Interact.* **19**(4), 319–349 (2004)
- Hassenzahl, M., Monk, A.: The inference of perceived usability from beauty. *Hum. Comput. Interact.* **25**(3), 235–260 (2010)
- Holmqvist, K., Nyström, M., Mulvey, F.: Eye tracker data quality: what it is and how to measure it. In: Proceedings of the Symposium on Eye Tracking Research and Applications, pp. 45–52. ACM (2012)
- Imberger, W., Schrepp, M., Held, T.: Cognitive processes causing the relationship between aesthetics and usability. In: Holzinger, A. (ed.) USAB 2008. LNCS, vol. 5298, pp. 43–54. Springer, Heidelberg (2008). doi:[10.1007/978-3-540-89350-9_4](https://doi.org/10.1007/978-3-540-89350-9_4)
- Kurosu, M., Kashimura, K.: Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. In: CHI 1995: Conference Companion on Human Factors in Computing Systems, pp. 292–293. ACM (1995)
- Lavie, T., Tractinsky, N.: Assessing dimensions of perceived visual aesthetics of web sites. *Int. J. Hum. Comput. Stud.* **60**(3), 269–298 (2004)
- Leuthold, S., Schmutz, P., Bargas-Avila, J.A., Tuch, A.N., Opwis, K.: Vertical versus dynamic menus on the world wide web: eye tracking study measuring the influence of menu design and task complexity on user performance and subjective preference. *Comput. Hum. Behav.* **27**(1), 459–472 (2011)
- Lindgaard, G.: Aesthetics, visual appeal, usability and user satisfaction: what do the user's eyes tell the user's brain? *Aust. J. Emerg. Technol. Soc.* **5**(1), 1–14 (2007)
- McCarthy, J.D., Sasse, M.A., Riegelsberger, J.: Could i have the menu please? An eye tracking study of design conventions. *People and Computers XVII — Designing for Society*, pp. 401–414. Springer, London (2004)
- Moshagen, M., Musch, J., Göritz, A.S.: A blessing, not a curse: experimental evidence for beneficial effects of visual aesthetics on performance. *Ergonomics* **52**(10), 1311–1320 (2009)
- Ngo, D., Samsudin, A., Abdullah, R.: Aesthetic measures for assessing graphic screens. *J. Inf. Sci. Eng.* **16**(1), 97–116 (2000)
- Ngo, D., Byrne, J.: Application of an aesthetic evaluation model to data entry screens. *Comput. Hum. Behav.* **17**(2), 149–185 (2001)
- Ngo, D.: Measuring the aesthetic elements of screen designs. *Displays* **22**(3), 73–78 (2001)
- Ngo, D., Teo, L., Byrne, J.: Modelling interface aesthetics. *Inf. Sci.* **152**, 25–46 (2003)
- Nielsen, J.: Heuristic evaluation. In: Nielsen, J., Mack, R.L. (eds.) *Usability Inspection Methods*, pp. 22–62. Wiley, New York (1994)
- Nielsen, J.: F-shaped pattern for reading web content. *Jacob Nielsen's Alertbox* (2006). <http://www.nngroup.com/articles/f-shaped-pattern-reading-web-content/>
- Outing, S., Rual, L.: The Best of Eyetrack III: What We Saw When We Looked Through Their Eyes. Poynter Institute (not dated). <http://poynterextra.org/eyetrack2004/main.htm>
- Phillips, C., Chaparro, B.: Visual appeal vs. usability: which one influences user perceptions of a website more. *Usability News* **11**(2), 1–9 (2009)
- Pieters, R., Wedel, M., Batra, R.: The stopping power of advertising: measures and effects of visual complexity. *J. Mark.* **74**(5), 48–60 (2010)
- Poole, A., Ball, L.J., Phillips, P.: In search of salience: a response time and eye movement analysis of bookmark recognition. In: Fincher, S., Markopolous, P., Moore, D., Ruddle, R. (eds.) *People and Computers XVIII-Design for Life: Proceedings of HCI 2004*. Springer, London (2004)

- Purchase, H.C., Hamer, J., Jamieson, A., Ryan, O.: Investigating objective measures of web page aesthetics and usability. In: Proceedings of the Twelfth Australasian User Interface Conference, vol. 117, pp. 19–28. Australian Computer Society (2011)
- Quinn, J.M., Tran, T.Q.: Attractive phones don't have to work better: independent effects of attractiveness, effectiveness, and efficiency on perceived usability. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 353–362. ACM (2010)
- Reinecke, K., Yeh, T., Miratrix, L., Mardiko, R., Zhao, Y., Liu, J., Gajos, K.Z.: Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2049–2058. ACM (2013)
- Rose, D.E., Levinson, D.: Understanding user goals in Web search. In: Proceedings of WWW 2004, pp. 13–19 (2004)
- Salvucci, D.D., Goldberg, J.H.: Identifying fixations and saccades in eye-tracking protocols. In: Proceedings of the 2000 Symposium on Eye Tracking Research & Applications, pp. 71–78. ACM (2000)
- Shenkman, B.O., Jönsson, F.: Aesthetics and preferences of web pages. *Behav. Inf. Technol.* **19**(5), 367–377 (2000)
- Sonderegger, A., Sauer, J.: The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Appl. Ergon.* **41**(3), 403–410 (2010)
- Thüring, M., Mahlke, S.: Usability, aesthetics and emotions in human–technology interaction. *Int. J. Psychol.* **42**(4), 253–264 (2007)
- Tractinsky, N.: Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In: CHI 1997: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 115–122. ACM (1997)
- Tractinsky, N., Katz, A.S., Ikar, D.: What is beautiful is usable. *Interact. Comput.* **13**(2), 127–145 (2000)
- Tuch, A.N., Roth, S.P., Hornbæk, K., Opwis, K., Bargas-Avila, J.A.: Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI. *Comput. Hum. Behav.* **28**(5), 1596–1607 (2012)
- Van Schaik, P., Ling, J.: The effect of link colour on information retrieval in educational intranet use. *Comput. Hum. Behav.* **19**(5), 553–564 (2003)
- Van Schaik, P., Ling, J.: Modelling user experience with web sites: usability, hedonic value, beauty and goodness. *Interact. Comput.* **20**(3), 419–432 (2008)
- Velichkovskiy, B.M.: Research on cognitive functions and today's technologies [Issledovanie kognitivnykh funktsiy i sovremennye tekhnologii]. *Bull. Russ. Acad. Sci.* **80**(5–6), 440–446 (2010)
- Wang, Q., Yang, S., Liu, M., Cao, Z., Ma, Q.: An eye-tracking study of website complexity from cognitive load perspective. *Decis. Support Syst.* **62**, 1–10 (2014)
- Zaphiris, P., Ellis, R.D.: Website usability and content accessibility of the top USA Universities. In: WebNet, pp. 1380–1385 (2001)