Facets of website content

Meinald T. Thielsch[1] & Gerrit Hirschfeld[2]

[1]Department of Psychology, University of Münster, Germany, [2]Faculty of Business Management and Social Sciences, University of Applied Sciences Osnabrück, Germany

ABSTRACT

Content is of primary importance in the World Wide Web. In particular, subjective perceptions of content are known to influence a variety of user evaluations thereby altering attitudes and behavioral outcomes. Thus, it is essential that individually experienced facets of content can be adequately assessed. In a series of seven studies we create, validate, and benchmark a measure for users' subjective view on web content. In the first six studies, a total of 3,106 participants evaluated a sum of 60 websites. The resulting Web-CLIC questionnaire is a 12-item measure based on prior research on web content. It encloses four main facets of users' content experience: *clarity*, *likeability*, *informativeness*, and *credibility* – jointly representing a general factor *subjective content perception*. Very high internal consistencies and high short- to medium-term retest reliabilities are demonstrated. Strong evidence for construct validity in terms of factorial, convergent, divergent, discriminative, concurrent, experimental, and predictive validity is found. In a seventh study, encompassing 7,379 ratings on 120 websites, benchmarks for ten different content domains and optimal cut points are provided. Overall, the present research suggests, that the Web-CLIC is a sound measure of subjective content perception of both practical and theoretical benefit.

CONTENTS

1. INTRODUCTION
    1.1. Related work
        Approaches to website content
        Assessment of subjective perceptions of web content
    1.2. Aims of the present study
2. STUDY 1. TESTING THE INITIAL ITEM SET AND EXPLORATORY FACTOR ANALYSIS
    2.1. Method
        Participants
        Stimulus material and measures
        Procedure
    2.2. Results and discussion
        Item characteristics
        Exploratory factor analysis
3. STUDY 2. CONFIRMATORY FACTOR ANALYSIS
    3.1. Method
        Participants
        Stimulus material and measures
        Procedure
    3.2. Results and discussion
4. STUDY 3. RELIABILITY OF THE WEB-CLIC
    4.1. Method
        Participants
        Stimulus material and measures
        Procedure
    4.2. Results and discussion
        Internal consistency
        Retest reliability
5. STUDY 4. CONSTRUCT VALIDATION OF THE WEB-CLIC
    5.1. Method
        Participants, stimulus material, and measures
        Procedure
    5.2. Results and discussion
6. STUDY 5. EXPERIMENTAL VALIDATION
    6.1. Method
        Participants
        Stimulus material and measures
        Procedure
    6.2. Results and discussion
        Manipulation check
        Experimental validation

7. STUDY 6. FURTHER VALIDATION AND USEFULNESS OF THE WEB-CLIC
7.1. Method
Participants
Stimulus material and measures
Procedure
7.2. Results and discussion
8. STUDY 7. BENCHMARKS AND OPTIMAL CUT POINTS FOR THE WEB-CLIC
8.1. Method
Participants
Stimulus material and procedure
8.2. Results and discussion
Influences of age, gender, and education level
Benchmarks for different website categories
Cut point analyses
9. GENERAL DISCUSSION
9.1. Facets of users' website content perception
9.2. Objectivity, reliability, and validity of the Web-CLIC
9.3. Interpretation of the Web-CLIC and practical implications
9.4. Limitations and future research
9.5. Conclusion
APPENDIX A. OVERVIEW ON FACETS OF WEBSITE CONTENT IN USER EXPERIENCE
APPENDIX B: ITEMS AND ITEM STATISTICS
B.1 Original items analyzed in study 1.
B.2 Scree plot resulting from exploratory factor analysis of study 1.
B.3. Items, factor loadings, means, and standard deviations for remaining 37 items in exploratory factor analysis of study 1.
APPENDIX C: STIMULI
C.1 Experts ratings for websites in pre-study of study 1
C.2 URLs of websites tested in study 2, 3, and 4.
C.3 Examples for treatments and stimuli used in study 5
APPENDIX D: BENCHMARK AND CUT POINT ANALYSES
D.1 Correlations between age, educational level and the Web-CLIC scores
D.2 ROC curve for the Web-CLIC against the dichotomous good versus unattractive rating

1. INTRODUCTION

"Content is king" (Fillmore, 1995).

The World Wide Web has become a constant companion in our daily life. Most of the time, we use it to search and receive specific pieces of information (Dinet, Chevalier, & Tricot, 2012; Koch & Frees, 2016). How users perceive the presented information, i.e. web content, is a primary factor for website success (Agarwal & Venkatesh, 2002; Palmer, 2002; Thielsch, Blotenberg & Jaron, 2014). There are several measures to investigate web users' impressions of usability and aesthetics – yet there is a lack of a standardized measure of web content perceptions. Imagine you are responsible for an e-health website aimed at helping people to stop smoking. To ensure maximum possible effectiveness of your website, you want it to be usable and pleasantly designed – and you will have no problems finding high quality instruments to test both of these aspects from the users' perspective. But, the most important part of this specific website is the content. Only if readers understand, believe, and appreciate the presented information, they are able and willing to use it, possibly leading to a higher chance to stop smoking (see Lehto & Oinas-Kukkonen, 2011). Yet, you will have major problems finding a practicable measure, that is reliable, specifically tailored to assess user's perceptions of web content, and adequately validated. The reason is that content is mostly considered only as a partial aspect in instruments aimed at website quality in general – or just tested with unidimensional single items and unaudited ad hoc scales (see below).

Thus, the aim of the present paper is to develop a questionnaire that assesses users' subjective perceptions of website content. Such a measure can help researchers and practitioners to a) improve the understanding of a website contents impact on users' behaviors, b) optimize websites for specific target groups and deliver best services possible, and c) analyze the interplay among content, usability and design evaluations. We define subjective perceptions of web content as users' general perceptions, impressions, and evaluations resulting from the interaction with presented content objects of a website. Based upon current theories of users' processing of websites (such as aesthetic perceptions, see Moshagen and Thielsch, 2010), we adopt an interactionist perspective: The formation of subjective perceptions relies on the interaction between characteristics of the perceiver, the use scenario, and properties of web content objects (as defined in ISO 9241-151; ISO, 2006). We concentrate on those facets that are best assessed using a survey approach and can be rated by typical users. In the following, we review current approaches to website content, its' subjective perceptions and previous measures, before describing a series of seven studies in which we develop and validate a novel instrument to assess the clarity, likeability, informativeness, and credibility of websites, called Web-CLIC.

1.1. Related work

Any typical corporate, institutional or private website is built to present specific information. Thus, there is a wide range of related research that aims to quantify different aspects of web content.

Approaches to website content

ISO 9241-151 defines content as “a set of content objects”, and content object as “interactive or non-interactive object containing information represented by text, image, video, sound or other types of media” (ISO, 2006, p. 3). In line with this technical description of content, a large body of research tries to extract measures of website quality and reputation from features such as key words, links, or syntactical structure. For example, several metrics, such as HITS (Kleinberg, 1999) or PageRank (Brin et al., 1998), attempt to analyze and rank websites based on link structure. Other metrics, such as BM25F (Robertson & Zaragoza, 2009), RankNet (Burges et al., 2005) or SocialPageRank (Bao et al., 2007), use query terms and the textual content of websites. Content objects and structures are used for automatic classification tasks (e.g., Cai et al., 2003; Dumais & Chen, 2000) and automatic content analysis (e.g., Kohli, Kaur & Singh, 2012; Serrano-Guerrero et al., 2015). These lines of research resulted in powerful classification and search tools. Yet, the content features of a website are perceived and interpreted by its users only. For example, an article on a specific disease may be deemed easy to read by experts in the field but unintelligible by others. Simply measuring syntactic properties or word-frequency, neglects interindividual differences that are important for comprehension and consequently for users’ appreciation of web content. Thus, in line with research on data quality (Wang & Strong, 1996) and information quality (Delone & McLean, 2003), websites are seen as information products for which subjective parameters should be evaluated (Wang et al., 1998).

From this perspective, perceptions of content need to be separated from perceptions of a websites’ design aesthetics[1] or usability[2]. Even though there are important relations between these constructs (see Thielsch et al., 2014), they can also be differentiated by the processes and time-scales at which they are formed: While aesthetic perceptions to a large degree are driven by the bottom-up processes of the human visual perception, perceptions of content are based on top-down processes, including reflective cognitive processes and reasoning (Dinet et al., 2012; Douneva, Jaron & Thielsch, 2016; Thielsch & Hirschfeld, 2012). Judgements about website aesthetics are built within a few hundred milliseconds (Bölte et al., 2017), while users need about three to four seconds to give first impression ratings about content credibility (Robins & Holmes, 2008). Thus, the processing of content and the processing of aesthetics are probably relying on different modules in the human brain, working at different time scales as well. Additionally, meaningful ratings of usability require even more time and users’ interaction with a website (Thielsch, Engel & Hirschfeld, 2015), while content ratings can be based on reading a few or even one webpage only. Most importantly, even when usability and aesthetics are perfectly optimized, users still might neglect a website when content is

[1] Website aesthetics is defined as ‘an immediate pleasurable subjective experience that is directed toward an object and not mediated by intervening reasoning’ (Moshagen and Thielsch, 2010, p. 690)

[2] Usability is defined as the ‘extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use’ (ISO, 1998, p. 2).

perceived as poor (e.g., Sillence et al., 2007). Finally, it is important to develop measures for subjective perceptions of content in the online context. While content is also important for offline media, for example newspapers and magazines, content perception online is different in that attention spans are rather short (e.g., Liu, White, & Dumais, 2010), while hypertext requires higher reading skills (e.g., Coiro, 2011). Moreover, consumers are much less committed to a single online source and can easily use search engines to access alternative content (Dinet et al., 2012).

Existing research stress the importance of user perceptions: A number of studies find that subjective perceptions of website content are systematically related to general user reactions, such as overall attitudes and satisfaction (e.g., Kang & Kim, 2006; Palmer, 2002; Shukla, Sharma, & Swami, 2010), perceived ease of use, usefulness, and usability (e.g., Ahn, Ryu, & Han, 2007; Thielsch et al., 2014), trust (e.g., De Wulf et al., 2006; Rahimnia & Hassanzadeh, 2013; Seckler et al., 2015), perceived website quality (e.g., Aladwani & Palvia, 2002; Kincl & Štrach, 2012), perceived overall service quality (e.g., Liu & Arnett, 2000; Yang et al., 2005), purchase intentions and sales performance (e.g., Hsieh et al., 2015; Shukla et al., 2010; Thongpapanl & Ashraf, 2011; Verhagen, Boter & Adelaar, 2010), website success, or website preference in terms of commitment, loyalty, and the intention to revisit (e.g., Aranyi & van Schaik, 2016; De Wulf et al., 2006; Kim & Niehm, 2009) or to recommend a website (e.g., Cober et al., 2003; Kim & Niehm, 2009; Thielsch et al., 2014). However, these studies are mostly correlational and use partly diverging conceptualizations of web content. As a result, neither the processes that give rise to these individual findings, the potential overlaps between content facets, nor their relation to perceptions of aesthetics and usability are sufficiently known. Still, the multitude of existing findings illustrates the importance of the subjective perception of web content and its' potential effects on actual behavior. Different strategies have been applied to examine web users content perceptions, as we will illustrate in the next section.

Assessment of subjective perceptions of web content

In research, five different strategies to assess subjective perceptions of content can be found: (1) single item assessments (partly enclosed in general website evaluation scales), (2) attribute lists and checklists, (3) unidimensional scales, (4) multidimensional scales enclosed in extensive measures of “website quality”, or (5) specific instruments designed to assess perceptions of website content.

(1) Some studies measure content perceptions with single items (e.g., Kincl & Štrach, 2012), or single items enclosed in general measurements of website perceptions (e.g., Karreman, van der Geest, & Buursink, 2007; Loiacono, Watson, & Goodhue, 2007; Liu & Arnett, 2000). However, single-item measures are not as reliable as multiple-item scales (Schmidt & Hunter, 1996; Spector, 1992) and therefore not well suited for the assessment of complex constructs (Baumgartner & Homburg, 1996).

(2) Other studies use attribute lists or checklists to evaluate website content (e.g., Agarwal & Venkatesh, 2002; Caro et al., 2008; Hasan & Abuelrub, 2011; Huizingh, 2000; Smith, 2001; Sutherland et al., 2005; Tsakonas & Papatheodorou, 2006), but some of those were constructed only for experts or webmasters. Additionally, while checklists

are well suited for an inventory or the assessment of frequencies of specific aspects, user perceptions of content aspects are difficult to assess with such an approach.

(3) Several authors use unidimensional scales for measuring subjective perceptions of content (e.g., Ahn et al., 2007; Cao, Zhang & Seydel, 2005; Geißler, Donath, & Jaron, 2003; Hausman & Siepke, 2009; Hong & Kim, 2004; Lin, 2007; Rahimnia & Hassanzadeh, 2013; Ranganathan & Ganapathy, 2002; Shukla et al., 2010). Here, content often is a subdomain only within more comprehensive questionnaires that aim to assess overall website quality, usability, or user experience. The major drawback of unidimensional scales is that they are based on the idea that it is impossible (or not necessary) to discern different facets of content perceptions. However, several authors suggest that there are multiple facets of content perceptions (e.g., Agarwal & Venkatesh, 2002; McKinney, Yoon & Zahedi, 2002; Yang et al., 2005). In practice, it may be more helpful to get more specific feedback about a website than just one unidimensional score. Additionally, many single-item and unidimensional scales lack a proper psychometric examination encompassing reliability, and validity analyses. Reliability sets an upper limit to the magnitude of relationships to other constructs. Validity, among other things, ensures that items are not confounded with other constructs. For example, the item "I can find what I need in the website" from the information quality scale of Cao et al. (2005), designed to grasp information relevance, might be influenced by usability issues.

(4) Multidimensional scales assessing different facets of website content are sometimes part of broad instruments measuring general attitudes towards a website (e.g., Abdinnour-Helm, Chaparro & Farmer, 2005; Aladwani, 2002; Aladwani & Palvia, 2002; Chakraborty, Srivastava, & Warren, 2005; De Wulf et al., 2006; Elling, Lentz & de Jong, 2007; Hong, 2006; Kang & Kim, 2006; McKinney et al., 2002). Again, only little information about the psychometric quality of these scales is available. Some studies inspect the factorial structures, but profound and systematic validations are missing.

(5) To the best of our knowledge, only two standardized instruments are published that are constructed with the sole purpose to assess users' subjective perceptions of website content: the ICTQ (Ozok & Salvendy, 2001) and the WWI (Thielsch, 2008). ICTQ stands for "**I**nterface **C**onsistency **T**esting **Q**uestionnaire", a measure consisting of 94 items on nine scales, addressing the consistency of text structure, general text features, information representation, lexical categories, meaning, user knowledge, text content, communicational attributes, and physical attributes (see Ozok & Salvendy, 2001). An original item set of 125 items, generated based on the literature, was reduced with a sample of 120 students via factor analysis and factor loadings as selection criteria. The internal consistency of the whole questionnaire was $\alpha = 0.81$, ranging from .79 to .85 for five of the nine subscales, while for four scales values were not available. The inter-rater reliability was 0.75 for the whole questionnaire, ranging from .68 to .82 for the scales. Furthermore, Ozok and Salvendy (2001) report an analysis with additional 20 engineering students and found mostly no differences in ICTQ factor scores between different student groups.
The WWI (in German "Fragebogen zur **W**ahrnehmung von **W**ebsite-**I**nhalten" [perception of website content questionnaire]) was created based on a literature search and a series of two studies (see Thielsch, 2008). Items were derived from existing scales

in the field, from market research, or were newly created. The initial item pool was evaluated and extended by 25 experts and 16 web users, then tested with N = 322 web users in a second study. Thielsch (2008) deleted items if there were floor or ceiling effects, bimodal answer distributions, or more than 10 % of participants indicating problems answering them. Remaining items were analyzed with an exploratory factor analysis, resulting in three factors explaining 54.40 % of the variance. The final version of the WWI was created considering factor loadings, item selectivity, specific contents of the items, and by using the tool "Alphamax" (Hayes, 2005). This led to three scales with three items each: "Liking" ($\alpha = .90$), "Intelligibility" ($\alpha = .78$), and "Quality and use" ($\alpha = .71$). Thielsch (2008) argues for objectivity in a web-based research scenario as well as for content validity due to the inductive and expert based construction and high correlation ($.92 \leq r \leq .95$) between full and reduced item sets of each scale.

Both, ICTQ and WWI, suffer from several shortcomings: First, Ozok and Salvendy (2001) used a relatively small sample for factor analyses of the ICTQ items (N = 120). The sample of Thielsch (2008) with N = 322 is better suited for this kind of analysis, but recent research suggests that one might need at least sample sizes of 500 to 1000 to find optimal item configurations in exploratory factor analysis (see Hirschfeld, von Brachel & Thielsch, 2014). Second, Cronbach's alphas for some scales are only satisfactory, or in case of the ICTQ, partly not available. Third, stability and retest reliability of both measures have not been tested so far. Fourth, an extensive validation is missing for ICTQ and WWI, including at least confirmatory analysis as well as convergent and divergent validation strategies. Fifth, the ICTQ has a very narrow focus on the consistency of website content, likewise there are important subjective content facets that were not tested in the construction of the WWI. Finally, from a practitioner's point of view, interpretation aids such as benchmarks are essential when using such a measure, but are not included in the ICTQ or WWI. Thus, from our point of view, a standardized, fully proved, validated and practical measure to assess subjective perceptions of web content is still lacking.

1.2. Aims of the present study

The aim of the present research is to create, validate, and benchmark a sound measure of subjective website content perception. Based on a literature search and on existing instruments (especially the WWI; Thielsch, 2008), we aim to create an empirically supported measure that is short and thus easy to apply in different evaluation settings. Therefore, we identified most relevant facets of users' web content evaluation and compiled them together in one measure. This newly created instrument is tested with item analysis as well as with exploratory factor analysis (study 1), to determine which facets are indeed independent from one another and which ones can be merged. Focusing on only those scales that assess a unique factor, results in a short measure (especially if compared to a mix of the few available validated scales). We further verify this measure in confirmatory factor analysis (study 2). In addition to a thorough inspection of the classical psychometric quality criteria reliability (study 3) and validity (study 4, 5, and 6), we give advice for interpretation and practical use by providing benchmarks as well as optimal cut points (study 7). For an overview of study aims and methods see Figure 1.

Figure 1: Aims and methods of studies

Study	Aim	Method
Study 1	Test the initial item set and explore factor structure	Descriptive analysis of item characteristics and exploratory factor analysis
Study 2	Replicate factor structure found in study 1 (including a general factor)	Confirmatory factor analysis (based on a sample different to study 1)
Study 3	Analyses of reliability (internal consistency) and test-retest reliability	Cronbach's α and test-retest correlations
Study 4	Construct validation: Investigation of convergent, divergent, concurrent, and discriminative validity	Correlations with related, unrelated, and simultaneously assessed criteria; MANOVA (analyzing Web-CLIC scores on different websites)
Study 5	Experimental validation: Testing the sensitivity of scales to corresponding changes in website content	Systematic variation of a test website; MANOVA (analyzing Web-CLIC scores as a function of website changes)
Study 6	Testing the usefulness of the Web-CLIC by 1) a comparison of with global ratings, 2) analyzing task dependency, 3) predicting behavioral intentions and actual behavior	Experimental study systematically varying tasks while asking for user evaluations, behavior intentions, and actual decision behavior
Study 7	Providing guidelines for practical application	Analyses of benchmarks and optimal cut points

2. STUDY 1. TESTING THE INITIAL ITEM SET AND EXPLORATORY FACTOR ANALYSIS

Aim of study 1 was to explore the factors that underlie different items designed to capture diverse facets of subjective perceptions of website content, and to reduce a large item pool based on the prior research. The initial item set should only contain items representing facets of subjective perceptions of website content that can be best rated by typical users. Some facets that are often assessed, such as availability, amount of information or security, may be evaluated via user ratings – but automatic, algorithm-based measures will be better suited or could be performed quicker. Other, in study 1 excluded, content facets might be well assessable in expert studies but not so much in regular user evaluations: For example, facets such as completeness, originality or timeliness will require specific knowledge for a sound assessment. Thus, we focus on the content facets best suited for a survey approach as presented in Appendix A (with the exception of the facet *perception of specific content* as we aimed for a universal evaluation instrument). We collected a set of 40 items (see Appendix B.1, for the full item pool including references): The facet *clarity/comprehensibility* is represented by seven items, *credibility* by eight items, *informativeness* by five items, *likeability/attractiveness* by six items, *relevance* by five items, *originality/uniqueness of content* by four items, and *usefulness* by five items. Those 40 items were taken or adapted from prior measures of website content; in particular including all nine items of the WWI as well as additional items of its draft version (Thielsch, 2008). All items were revised in respect of wording before they were tested in study 1.

2.1. Method

Participants

A total of 1,226 participants took part in this web-based study; 698 were female (56.9 %), 528 male (43.1 %). Ages ranged from 14 to 67 years (*M* = 23.15, *SD* = 3.56). The education level of about 95.6 % of the participants was Abitur (German university entrance qualification) or higher. On average, the participants had been using the Internet for 9.28 years (*Min* = 2, *Max* = 26, *SD* = 2.60) and stated an active use of on average 2.83 hours a day (*Min* = 1, *Max* = 14, *SD* = 1.89). Participants took part voluntarily and on an anonymous basis without any compensation.

Stimulus material and measures

A pre-study was performed to pre-select a stimulus set unknown to participants but still reflecting a typical range in general website content quality. Therefore, N = 37 experts (12 female, 25 male) were recruited at the end of October 2010 via the German Internet Research List (gir-l) and an online forum of the German UPA (German Usability and User Experience Professionals Association). Experts were working in the area of online research, usability consulting, and web content creation; mean age was 37.81 years (*SD* = 7.26), average Internet experience 14.43 years (*SD* = 3.00). The experts randomly rated 19 websites from six different content domains (see Appendix C.1; screenshots can be requested via the corresponding author) on a seven-point Likert scale with respect to content quality, dichotomously for level of familiarity (known/unknown), as well as on a six-point grading scale with respect to the overall impression. Ten websites were selected, representing a maximal possible range of content quality with an even distribution of websites within this range. Additionally, only mostly unfamiliar websites, with expert evaluations that were not influenced by age or gender, were selected for the final set (see Appendix C.1).

The initial pool of 40 items (as described above and in Appendix B.1) was used to define the first version of the newly created instrument. All items were scaled on a seven-point Likert scale ranging from 1 (“strongly disagree”) to 7 (“strongly agree”).

Procedure

Participants were recruited via social networks, using a mailing list of the German National Academic Foundation, and at the Department of Psychology at the University of Münster. Participants were informed about objective, principle investigator, anonymity, voluntariness and duration of the present study. After being asked for some demographic information (e.g., age, gender, education level, Internet experience), participants were randomly assigned to one fully functional website from the stimulus set. The website in question was presented within a split screen, the items were presented in a smaller upper panel. At the beginning, participants were asked to rate their first impression of the website. Next, they were instructed to explore the given website and to open some subpages (i.e., the task was free exploration). Then, they answered the 40 items regarding content quality (and four other measures, see study 4). The items and scales used in this part of the study were given in random order. Additionally, the overall impression and the intention to revisit the website were rated. At the end of the study, participants were thanked. They were given the opportunity to exclude their data from the subsequent analysis and to comment on the study. The study was available online from 11/23/2010 till 12/07/2010; on average participants needed 15 minutes to complete.

2.2. Results and discussion

Item characteristics

In a first step, we used item analysis to exclude items with extreme skew and/or difficulty. The distribution of responses was extremely skewed for three items (07, 16, 35; see Appendix B.1 for item wordings and source) and these were excluded from further analysis. The remaining items had levels of skewness ($-0.895 \leq skew \leq 0.886$) and kurtosis ($-1.054 \leq kurtosis \leq 0.385$) that are acceptable for factor analysis (see West, Finch & Curran, 1995).

Exploratory factor analysis

In a second step, we performed an exploratory factor analysis on the remaining 37 items to determine the factors and select items, following the recommendations by Costello and Osborne (2005). Specifically, we used factor analysis with oblique rotation to extract factors. The number of factors was determined based on the scree plot and an inspection of the resulting loading pattern. For the loading patterns, we required that all retained factors should have at least three items, which only show substantial loadings (>0.3) on the respective factor and no substantial cross-loadings, i.e. simple structure (Costello & Osborne, 2005). Based on the scree plot (see Appendix B.2), different numbers of factors were extracted (2,3,4, and 5). Of these, the solution with four factors explained 51% of the variance and yielded a loading pattern that could be readily interpreted. Extracting five factors resulted in a solution in which all items that loaded on the fifth factor also had strong cross-loadings on other factors. Furthermore, when extracting less than four factors items are lumped together belonging to separate facets of subjective content perceptions: In the two-factor solution, items from the clarity, informativeness, and credibility factors lump together, and the second factor encompass items related to likeability. In the three-factor solution, items from the informativeness and credibility factors lump together, and likeability and clarity form two separate factors. In the preferred four factor solution, the first factor, *likeability,* comprised eleven items that are all concerning the general positive evaluation of the website content, e.g. “I enjoy reading the website”. The second factor, *credibility,* comprised eight items, all indicating whether or not participants perceived the websites content as trustworthy or unbiased, e.g. “I can trust the information on the website”. The third factor, *clarity,* comprised nine items related to the way the information is presented and summarized, e.g. “The language used in the texts is current and easy to understand”. The fourth factor, *informativeness,* comprised nine items, all related to the potential value of the information that was presented, e.g. “The website is informative”. Thus, of the seven facets of subjective web content perception on which the items were based on, four were directly represented as factors, while the facets *relevance*, *usefulness*, and *originality/uniqueness of content* did not emerge as separate factors. Especially, four of five items that were supposed to assess *relevance* showed cross-loadings and thus were not included in the final questionnaire. Still, the facet relevance could be of importance in specific situations, especially when users are personally affected (e.g., when visiting e-health websites). Readers interested in this facet are referred to the according scales provided by Cao and colleagues (2005), respectively Lee and colleagues (2002). Items belonging to the *originality/uniqueness* facet showed strong and specific loadings on the likeability factor. Items from the *usefulness* facet loaded on the factors *likeability*, *clarity* and *informativeness*. Of these, two items were included in the final questionnaire in *clarity* and *informativeness*, because they reflected the breadth of these constructs.

Items were selected for inclusion in the final item-set based on (1) simple structure, and (2) meaning (Costello & Osborne, 2005). This led to the direct selection of seven items (number 21 and 25 for the likability factor, number 11 and 12 for credibility, number 02 for clarity, and number 19 and 36 for informativeness). Five additional items were selected, because they reflected different aspects of the supposed factor while still showing substantial loadings (see Appendix B.3). In doing so, items were preferred a) if they were empirically proven in several other studies and validated questionnaires (that is why item 24 was preferred instead of item 33 for likeability, item 17 instead of number 20 for informativeness, and item 37 instead of 05 for clarity), and b) if they were better worded in terms of being more common and easier to understand (that is why item 14 was preferred instead of item 13 for credibility), and focused on broad aspects (leading to a preference for item 04 instead of item 03 for the facet clarity). Thus, for each of the four factors, it was possible to select three items reflecting the specific content and conformed to simple structure. Only one item (number 17, “The information is of high quality.”) was selected for the factor *informativeness* even though it showed a cross-loading (of 0.306) on the factor *credibility*. This was done, because perceived quality of the presented information was deemed theoretically important, based on prior research on this aspect (Cao et al., 2005; Kim & Lim, 2001; Thielsch, 2008). The items that were finally selected are displayed in Figure 3.

The intercorrelations among means of the four scales ranged from .40 (likeability with credibility) to .71 (credibility with informativeness), indicating a possible overlap between these facets for the full item set (see Figure 4). We believe that these intercorrelations may be best explained by a general factor, that indicates positive evaluation of the website content, and thus tested for a g-factor structure in study 2. As there is only little evidence on the psychometric properties of items designed to capture various facets of subjective web content perceptions, we can only speculate why only some of the various facets put forward in the literature emerged as unique factors. It seems that informativeness and credibility are most similar, while clarity is more separate, and likeability the aspect that can be discerned most easily. This is in line with recent research and the idea that the quality of information is used as a cue for credibility (e.g., Appelman & Sundar, 2016; Metzger & Flanagin, 2013). At the same time, informativeness is often treated as a separate facet of content perception (see Appendix A), and correlations between scales are not as high as that the scales have to be joined (see Figure 4).

3. STUDY 2. CONFIRMATORY FACTOR ANALYSIS

Aim of study 2 is to replicate the factor structure found in study 1, additionally including a general factor in a confirmatory factor analysis (CFA). Therefore, we reanalyzed a data set of Hirschfeld and Thielsch (2015), which was so far only partly used for finding optimal cut points for an aesthetics measure. Up to now, those data had not been analyzed with respect to website content.

3.1. Method

Participants

A total of 618 participants took part in this web-based study; 321 were female (51.9 %), 297 male (48.1 %). Ages ranged from 15 to 82 years (*M* = 34.94, *SD* = 13.65). The education level of 78.7 % of the participants was Abitur (German university entrance qualification) or higher. On average, the participants had been using the Internet for 11.66 years (*Min* = 2, *Max* = 30, *SD* = 5.12) and stated an active use of on average 2.52 hours a day (*Min* = 0.2, *Max* = 12, *SD* = 1.92). Participants took part voluntarily and on an anonymous basis without any compensation.

Stimulus material and measures

A set of 30 websites from ten different content domains was used (information on the categorization scheme can be found in Thielsch, 2008; p. 86f. and in Appendix C.2; screenshots can be requested via the corresponding author). These websites were selected to represent a broad range of corporate and institutional websites in Germany, covering a huge percentage of a person's everyday life online activities. Each website category was represented by two to five websites (see Appendix C.2).
The twelve items identified in study 1 (see Figure 3) were used to define the final version of the instrument.

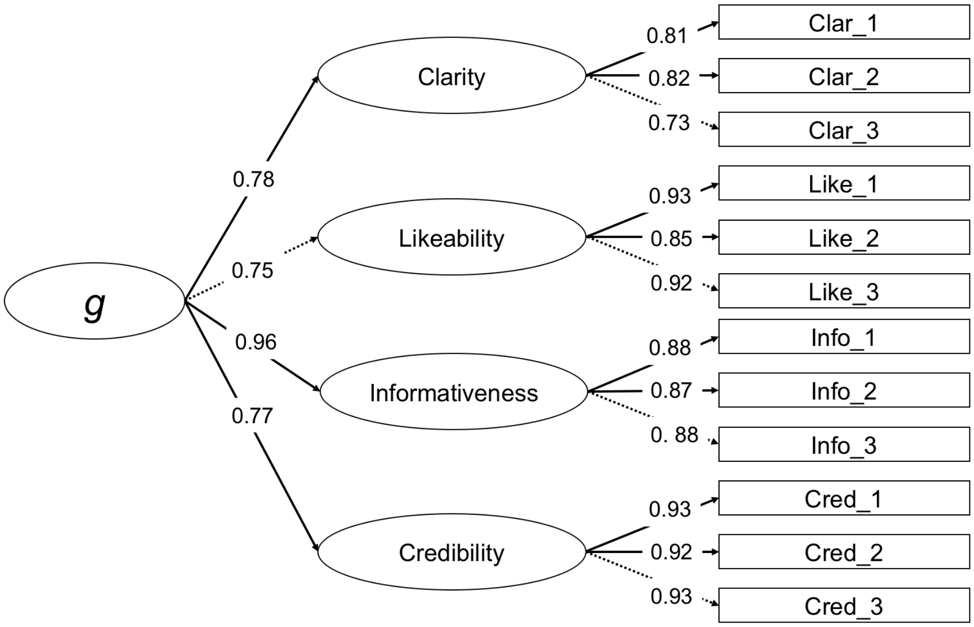
Procedure

Participants were recruited via the German online panel PsyWeb (https://psyweb.uni-muenster.de/). Participation in this panel is completely voluntarily and members agree on receiving invitations to scientific studies; they can unsubscribe and delete their personal data at any time. Participants of the present study received an e-mail inviting them to a study about the evaluation of websites. Following the invitation link, they were informed about involved researchers, anonymity, voluntariness and duration of the study. After being asked for some demographic information (age, gender, education level, Internet experience), participants were randomly assigned to one website from the stimulus set. The fully functional website in question was presented within a split screen, the items were presented in the smaller upper panel. First, participants were asked to rate their first impression of the website. Next, they were instructed to explore the given website and to open some subpages (i.e., the task was free exploration). Then, they answered the twelve content evaluation items identified in study 1 and two other measures (one for usability, one for aesthetics) not pertinent to this study. The measures used in the middle part of the study were given in random order, and all items within the questionnaires were also randomized. Afterwards, the overall impression was rated on the same scale as used at the beginning. At the end, participants could comment on the study, they were thanked and had the opportunity to exclude their data from the subsequent analysis. The study was available online from 10/30/2011 till 04/12/2012; participation on average took 10 to 12 minutes.

3.2. Results and discussion

A CFA was used to test the proposed structure of four factors with three items each and a second-order g-factor that had loadings on all four factors (see Figure 2). In order to estimate the model parameters, maximum likelihood estimation was used. Model fit was deemed acceptable if *CFI* and *TLI* >.95 and *RMSEA* < .08 (Hu & Bentler, 1999). The model fits the proposed structure very well as indexed by the various fit-indices (*CFI* = .98; *TLI* = .98; *RMSEA* = .058). All items showed large (at least .73) and statistically significant loadings on the proposed factors (see Figure 3). The g-factor also showed large loadings on the four factors (see Figure 3).

Figure 2. Structural model of the Web-CLIC.



Thus, we confirmed the proposed model with a general factor and four subscales: The *clarity* scale assesses how users perceive the intelligibility of web contents, the extent to which these are presented in a clear and concise manner, and the comprehensibility of the used language. The importance of an easy to understand content was already stressed as part of information quality in the Delone and McLean model (2003). Accordingly, aspects of clarity are enclosed in several other measures of website content (Aladwani & Palvia, 2002; De Marsico & Levialdi, 2004; Thielsch, 2008).

The *likeability* scale assesses users' perceptions of the attractiveness of a website regarding the content (not to be confused with attractiveness in terms of design aesthetics). Thus, on this scale the amount of interest, excitement, and joy caused by a given content is indicated. The importance of those aspects has also been stressed in prior research (e.g., Caro et al., 2008; Huizingh, 2000), and they are enclosed in some existing instruments (e.g., Kang & Kim, 2006; Thielsch, 2008).

The *informativeness* scale assesses the perceived amount of valuable and useful information given in a website. This facet of website content perception is enclosed in many existing measures of website content (e.g., Chakraborty et al., 2005; Hausman & Siepke, 2009; Kang & Kim, 2006; Lin, 2007; Shukla et al., 2010). The g-factor we found was most strongly related to the informativeness factor, indicating that this facet is central to the overall perception of website content. However, in different settings the relevance of the different facets may shift, for example credibility might be more important when banking or shopping websites are rated than it is when leisure websites are rated (see Casaló, Flavián, & Guinalíu, 2007).

Items selected for the *credibility* scale focus on aspects of authenticity, reliability, and trustworthiness of a given website content. Credibility is often focused in research and enclosed in many measures (e.g., Appelman & Sundar, 2016; De Wulf et al., 2006; Flanagin & Metzger, 2000; Fogg et al., 2001; Hong, 2006; Metzger, 2007; Wathen & Burkell, 2002). In the context of the Internet, credibility is described as believability of information and/or its source (e.g., Fogg & Tseng, 1999; Fogg et al., 2001), and as a receiver-based judgement with the two primary dimensions expertise and trustworthiness (see Metzger, 2007). Yet, the conceptualization and definition of credibility in digital communication is still under debate (see Metzger & Flanagin, 2013), and competing approaches can be found, such as the MAIN model (Sundar, 2008) or adoptions of the ABI-model of trust (Mayer, Davis, & Schoorman, 1995) on website credibility (e.g., Casaló et al., 2007; Flavián, Guinalíu & Gurrea, 2006). In contrast to these highly detailed conceptualizations of credibility, several researchers developed measures to evaluate website credibility on a global level (e.g., Choi & Rifon, 2002; De Wulf et al., 2006; Johnson & Kaye, 2002; Rains & Karmikel, 2009; Robins & Holmes, 2008), which mostly focus on aspects of message credibility rather than the credibility of the source (see Appelman & Sundar, 2016). This is in line with research identifying trustworthiness of information on a website as one of the most important criteria for website credibility (Warnick, 2004). The Web-CLIC credibility scale followed this general approach.

In conclusion, Study 2 confirmed the assumed structure of the instrument with four facets of subjective perceptions of website content representing one general factor. Thus, based on the scale names, this novel questionnaire was named Web-CLIC: **Web**site - **C**larity, **L**ikeability, **I**nformativeness, and **C**redibility.

Figure 3. Items selected in study 1 and loadings as found in study 2.

Item number and item		Factor			
		Clarity	Likeability	Informativeness	Credibility
			Item Loadings		
02	The contents of the website are clearly presented.	0.813			
37	The texts provide me information in a clear and concise manner.	0.817			
04	The language used in the texts is current and easy to understand.	0.729			
21	The website arouses my interest.		0.929		
25	The contents of the website are exciting.		0.853		
24	I enjoy reading the website.		0.917		
17	The information is of high quality.			0.879	
36	I find the information on the website to be useful.			0.865	
19	The website is informative.			0.878	
11	I find the information provided on the website to be authentic.				0.931
14	The information provided on the website is reliable.				0.92
12	I can trust the information on the website.				0.931
			Second-order loadings		
g		0.776	0.746	0.959	0.774

Figure 4: Intercorrelations among scale-means in study 1 and 2

	Clarity	Likeability	Informativeness	Credibility
Clarity	1	0.429	0.472	0.481
Likeability	0.602	1	0.579	0.400
Informativeness	0.625	0.653	1	0.708
Credibility	0.513	0.478	0,707	1

Note. All $p < .001$. Correlations displayed above the diagonal are from study 1 (N = 1226), below the diagonal from study 2 (N = 618).

4. STUDY 3. RELIABILITY OF THE WEB-CLIC

The aim of study 3 is to examine reliability (in terms of internal consistency) and test-retest reliability of the Web-CLIC. To analyze short-term and medium-term stability of the questionnaire, we conducted a study with three data collection points and time gaps between one day and two weeks between them.

4.1. Method

Participants

A total of 390 participants took part at the first measurement of this web-based study; 228 of them were female (58.5 %), 162 male (41.5 %). Ages ranged from 16 to 70 years (*M* = 45.22, *SD* = 14.02). The education level of 64.1 % of the participants was Abitur (German university entrance qualification) or higher. On average, the participants had been using the Internet for 14.98 years (*Min* = 3, *Max* = 34, *SD* = 4.63) and stated an active use of on average 2.34 hours a day (*Min* = 0.2, *Max* = 12, *SD* = 1.78). Participants took part voluntarily and on an anonymous basis; they had a chance to win one out of ten 10 € vouchers for an online bookshop. At the second time of measurement, n = 272 participants completed the study, n = 254 at the third.

Stimulus material and measures

Eight different websites served as stimulus material in this study (see Appendix C.2, screenshots can be requested from the corresponding author). Seven were chosen to cover a broad range of different website categories. The eighth website was a mock site with health-related medical and psychological information (named MedOnline), which had been created by an experienced web designer for research purposes. The websites were chosen under the guiding principle of prototypically, and ideally should not be known by the participants. In addition, content sum scores were supposed to show variance, so that floor or ceiling effects are prevented: The website evaluated worst (a download and software site) significantly scored lower on the Web-CLIC sum score than the website evaluated best (the information mock site), t (106) = -5.35, $p < .001$, $d = .87$[3]. Due to the dynamic nature of the Internet, the websites were monitored for the duration of the study. Only slight changes appeared between T1 and T3, as only on an information website (the homepage of a German newspaper) and on an e-commerce site content was edited on a daily basis. However, content domain and focus, writing style, layout, and general structure remained the same for both tested websites.
The final version of the Web-CLIC as identified in study 1 and 2 (see Figure 3) was used at all measurement dates.

[3] According to the guidelines provided by Cohen (1988), standardized mean differences of 0.2, 0.5 and 0.8 are considered small, medium, and large effects, respectively.

Procedure

Three data collection points were planned to measure the short-term (one day) and the medium-term (two weeks) stability of the Web-CLIC. The participants received an invitation for the first data point (T1) via e-mail, sent by the online panel PsyWeb on June 10, 2014. One day after T1, the invitation for the second time of measurement was sent (T2), and two weeks from T1 for the third one (T3). Every participant evaluated only one (at T1 randomly assigned) website at each time of measurement.

At T1, participants received information about objective, principal investigator, anonymity, voluntariness, the lottery of vouchers (for all participants completing T1, T2, and T3), duration, and design of the study. After consent was given, participants were asked for some demographical information (e.g., age, gender, education level, Internet experience). Then, as in our previous studies, the fully functional website and items were presented within a split screen. Participants were asked to complete a simple search-task in a depth of maximum two clicks without time limit (e.g., the task was searching for contact information for a telephone call). After that, the Web-CLIC scales (and three other measures regarding usability, recommendation, and aesthetics, all not pertinent to the present study) were presented in randomized order. Afterwards, the overall impression of the website in question was measured with four items. At the end of T1, participants again were asked for their consent, had the opportunity to exclude their data and to give additional comments. Participants needed about 10 minutes to complete T1.

At T2 and T3, a short introduction including a reminder about the study was given at each instance. The participants were asked for consent again; afterwards, Web-CLIC and additional measures were presented in the same way as in study 1, with the full-functional website displayed in a frame. At the end of each data collection, participants had the opportunity to exclude their data from subsequent analysis and to give additional comments. Participants on average needed about 5 minutes to complete each measurement. Additionally, at the end of T3, they were linked to a separated website (to guarantee anonymity) on which they could participate in the lottery.

4.2. Results and discussion

Internal consistency

Internal consistency is often considered as an indicator for reliability. Thus, we calculated Cronbach's α for the Web-CLIC scales and the sum score, based on the data gathered at T1 in the current study and based on the data of study 2 (see Figure 5). Given the guidelines of Nunnally (1978), Cronbach's α values above .8 can be considered as good, above .9 as excellent. For the Web-CLIC scales, Cronbach's α in both studies occurred above .8 ($.826 \leq \alpha \leq .949$), and above .9 for the sum score ($.920 \leq \alpha \leq .936$). Thus, the Web-CLIC exhibited good to excellent internal consistencies. Especially in the light of the shortness of the scales, each comprising only three items, those values are notable.

Figure 5: Internal consistencies (Cronbach's α) for Web-CLIC scales and sum score

	Web-CLIC sum score	Clarity	Likeability	Informativeness	Credibility
Study 2 (N = 618)	.936	.826	.927	.906	.949
Study 3, T1 (N = 390)	.920	.828	.922	.886	.934

Retest reliability

While the internal consistency can give an impression about homogeneity of a scale and accuracy of item configuration, retest reliability can be interpreted in terms of stability of a measure. For the calculation of short-term stability, the Web-CLIC again was given one day later, and, to test medium-term stability, also after two weeks. Retest reliability is interpreted under the light of the given time span between measures, values above .8 are considered as good, values above .7 as sufficient, and values above .6 as acceptable for research purposes and for analyses on group level (Nunnally, 1978). Results for the Web-CLIC are presented in Figure 6, showing sufficient to good retest values for the short-term stability of the Web-CLIC scales and sum score ($.779 \leq r_{T1\text{–}T2} \leq .892$). Likewise, sufficient to good retest values were found for the medium-term stability ($.713 \leq r_{T1\text{–}T3} \leq .836$), except for the clarity scale ($r_{T1\text{–}T3} = .688$).[4] Thus, the Web-CLIC appears to be a stable measure, at least over short and medium periods.

Figure 6: Short- and medium-term retest reliability for Web-CLIC scales and sum score

	Web-CLIC sum score	Clarity	Likeability	Informativeness	Credibility
Short-term stability (1 day, n = 272)	.892 (.865 - .914)	.779 (.727 - .882)	.819 (.776 - .855)	.781 (.730- .823)	.794(.745 - .834)
Medium-term stability (14 days, n = 254)	.836 (.795 - .870)	.688 (.617 - -748)	.812 (.765 - .850)	.713 (.647 - .769)	.773 (.718 - .818)

Note: All correlations are significant with $p < .001$; confidence intervals are given in parentheses.

[4] We conducted an additional forth measurement one year later in which n = 216 participants took part. Although we found significant retest correlations ($r_{T1\text{–}T4} = .636$ for the sum score, $.487 \leq r_{T1\text{–}T4} \leq .636$ for Web-CLIC scales), we decided not to report those results in detail, as we are not able to determine whether the decrease in correlations occurs due to aspects of users, websites, evaluated construct or the instrument. Thus, further research is needed to determine longitudinal effects in web content perception.

5. STUDY 4. CONSTRUCT VALIDATION OF THE WEB-CLIC

The purpose of study 4 is to validate the Web-CLIC using several validation strategies such as examining convergent validity (high correlations with related constructs), divergent validity (lower to no connections to unrelated criteria), discriminative validity (for the Web-CLIC the ability to distinguish between different websites), and concurrent validity (correlations to a simultaneously assessed criterion).

5.1. Method

Participants, stimulus material, and measures

Study 4 is based on the same sample and the same ten websites as described in study 1 (see Appendix C.2, screenshots can be requested via the corresponding author). For the construct validation of the Web-CLIC, several established measures were used. Unless otherwise specified, participants were asked to indicate their level of agreement to each item of these questionnaires on seven-point Likert scales ranging from 1 ("strongly disagree") to 7 ("strongly agree").

Informativeness and entertainment (Kang & Kim, 2006): Two single items from the main study of Kang and Kim (2006) were used (informativeness: "This web site is a valuable resource."; entertainment: „This web site is fun to explore."). Kang and Kim (2006) provided evidence for reliability and discriminant validity of their measure. In the current study, it is used as a criterion for convergent validity of the Web-CLIC overall sum score, as well as for the informativeness scale, and the likeability scale (where high correlations with entertainment were expected) respectively.

Overall impression of interestingness: Participants were asked to rate the overall interestingness of the given website with a single item ("Altogether, I think the content of this website is interesting"). This holistic item was used as criterion for convergent validity, especially for the Web-CLIC sum score.

Perceived website usability (PWU): This one-dimensional scale, measuring perceived website usability, was adapted to German based on Flavián et al. (2006). The PWU is a seven-item measure assessing perceived ease of use, ease of understanding and speed of information retrieval (see Thielsch, 2008; Thielsch et al., 2015). Thielsch (2008) found a Cronbach's α of .95 for the adapted version and provided evidence for factor and convergent validity. The PWU is used as a criterion for divergent validity.

Visual aesthetics of websites inventory (VisAWI): Moshagen and Thielsch (2010) created this 18 item-questionnaire to measure a general factor *subjective aesthetics* consisting of the four facets *simplicity*, *diversity*, *color*, and *craftsmanship*. The authors report Cronbach's α values between .85 and .94, and provided evidence for convergent, divergent, discriminative, concurrent and experimental validity. Additional analyses of the VisAWI can be found at Moshagen and Thielsch (2013), as well as at Hirschfeld and Thielsch (2015). The VisAWI is used as a criterion for divergent validity.

Overall website score: The overall website impression was assessed with a grade on a on a six-point grading scale (“Altogether: I would mark the website with…”, 1 = “very good”, 2 = “good”, 3 = “satisfactory”, 4 = “adequate”, 5 = poor 6 = “unsatisfactory”) commonly used in German education system. This grade was used as a criterion for concurrent validity.

Intention to revisit: The four items created for study five of Moshagen and Thielsch (2010) were used to assess participants’ intention to revisit the website (“I will visit the website again”, “I will visit the website on a regular basis”, “I would recommend the website to my friends”, “If I had interest in the content of the website in the future, I would consider visiting the website”). The responses to these items were averaged to form an index of the participants’ intentions to revisit the website. This index is used as a criterion for concurrent validity.

Procedure

The procedure was the same as described in study 1. Participants were informed about objective, responsible researchers, anonymity, voluntariness and duration of the study. After providing demographic information, participants were randomly assigned to one of the ten fully functional websites from the stimulus set (see Appendix C.2, screenshots can be requested via the corresponding author), and asked to browse the given website (i.e., free exploration task). As before, the given website and the items were presented within a split screen. Participants answered the items from the measure of Kang and Kim (2006), the Web-CLIC, the PWU, and the VisAWI. Items and questionnaires were presented fully randomized. Afterwards, the overall impression of interestingness, the overall website score, and the intention to revisit the website were rated. At the end, participants could exclude their data from the subsequent analyses, comment on the study, and were thanked. On average, they needed 15 minutes to complete the study.

5.2. Results and discussion

Correlations between the Web-CLIC and the convergent, divergent and concurrent criteria are shown in Figure 7. As expected, the Web-CLIC sum score showed high correlations with convergent criteria. In particular, high correlations were found between the Web-CLIC informativeness scale and the corresponding informativeness item of Kang and Kim (2006), as well as between the likeability scale and the entertainment item of Kang and Kim (2006). Other Web-CLIC scales correlated with those criteria to a lower extend. Sum score and likeability scale were highly correlated with the overall interestingness of a website, showing high agreement with a theoretically highly related holistic item.

Figure 7: Correlations between the Web-CLIC and the convergent, divergent and concurrent criteria

	WEB-CLIC sum score	Clarity	Likeability	Informative ness	Credibility
Convergent measures					
Informativeness	.731	.399	.560	.757	.594
Entertainment	.623	.440	.736	.467	.346
Overall impression: interestingness	.674	.362	.787	.581	.407
Divergent measures					
Perceived usability	.523	.622	.321	.361	.387
Perceived aesthetics	.545	.570	.433	.375	.378
Concurrent measures					
Overall website score	.707	.603	.571	.568	.520
Intention to revisit	.684	.448	.714	.577	.439

Note: N = 1226, all correlations are significant with $p < .001$. Overall website score was recoded so that high Web-CLIC values correspond to high overall scores.

Divergent validity refers to the degree to which the instrument is distinct from scales assessing other facets of subjective perceptions. Web-CLIC correlations to divergent constructs were lower (showing less connections to theoretically less related constructs) for most scales apart from the clarity scale and the sum score. The latter two, especially the clarity scale, showed high correlations with usability and aesthetics. This is in line with prior findings of such correlations (e.g., Aladwani & Palvia, 2002; Thielsch, 2008; Thielsch et al., 2014) and the interpretation of Moshagen and Thielsch (2010, p. 701) that good designers strive to jointly optimize content, usability and aesthetics. Particularly, clarity of website content can support usability (e.g., well-structured and comprehensible contents may help navigating the website), thus usability is not necessarily to treat as a divergent construct for this specific Web-CLIC facet. Still, in the light of such mixed results, additional analysis and an experimental validation of the Web-CLIC seem necessary (see study 5).

Web-CLIC correlations to concurrent measures were high (see Figure 7), especially between sum score and overall website score, as well as between the intention to revisit, sum score, and likeability scale. These results are in line with prior research, stressing the importance for website content perceptions for users' overall attitudes and satisfaction (e.g. De Wulf et al., 2006; McKinney et al., 2002; Shukla et al., 2010), their intention to revisit and loyalty (e.g. Aranyi & van Schaik, 2016; Kim & Niehm, 2009; Thielsch et al., 2014).

Finally, we analyzed the discriminative validity of the Web-CLIC, i.e. the ability of the measure to distinguish between different websites. To test whether the Web-CLIC sum score and scales differ as a function of the given website, a MANOVA was calculated (dependent variables: sum score and scales; independent variable: evaluated website). The overall MANOVA was significant, *F* (36, 4864) = 26.658, $p < .01$, $\eta^2 = .165$, indicating that websites received different evaluations on the Web-CLIC. Post-hoc univariate ANOVAs with website as independent variable and Web-CLIC sum score and scales as dependent variables showed significant differences for the sum score and all subscales ($12.648 \leq F\ (9, 1216) \leq 72.489$, all $p < 01$, $.086 \leq \eta^2 \leq .349$). In addition, when comparing the website evaluated most negatively (an entertainment website) with the one evaluated best (an e-recruiting website), a highly significant difference emerged (*t* (248) = 13.83, $p < .001$, $d = 1.75$), meaning that those two websites differ on the sum score by nearly two standard deviations. Thus, it can be concluded that the Web-CLIC is very capable of discriminating between different websites.

6. STUDY 5. EXPERIMENTAL VALIDATION

The purpose of study 5 is an experimental validation of the Web-CLIC. If validity is given, systematically manipulating the content of websites should significantly affect ratings on the Web-CLIC. Specifically, we manipulated several pages of a single website with respect to the website's clarity, informativeness, and credibility. We did not examine the likability facet, as manipulating it would have required an extensive study design including a pre-study examining users' personal web content interests and preferences with an exact matching in the following main study.

6.1. Method

Participants

A total of 567 participants took part in this study; 303 were female (53.4 %), 264 male (46.6 %). Ages ranged from 15 to 83 years ($M = 46.83$, $SD = 13.31$). The education level of 66.1 % of the participants was Abitur (German university entrance qualification) or higher. On average, the participants had been using the Internet for 13.71 years ($Min = 3$, $Max = 30$, $SD = 4.20$) and stated an active use of on average 1.79 hours a day ($Min = 0.02$, $Max = 15$, $SD = 1.84$). Participants took part voluntarily and on an anonymous basis; they received no compensation but could request a summary of the study's results.

Stimulus material and measures

We used a fully crossed 2x2x2 between-subject design. Subjects were randomly assigned to one of eight possible website versions. The website we manipulated was MedOnline, a fictional online portal created for experimental purposes providing health-related medical and psychological information for laypersons (it was already used in study 3 of the current paper). The experimental manipulation (see Appendix C.3) consisted of changes on five pages within the website.

Clarity was manipulated by changing features of the text. In the clearly intelligible condition, the text consisted of short sentences and avoided technical terms whenever possible. In the unclear condition texts consisted of long convoluted sentences (see Coleman, 1962) and many technical terms were used.

Informativeness was manipulated by changing the topics and information conveyed in the texts. In the low informativeness condition, the texts began with the topic mentioned in the headline but quickly drifted off to an entirely unrelated topic. Furthermore, the amount of useful information in these off-topic texts was limited, as only trivial information was provided. In the high informativeness condition, consistent and useful information were given.

Credibility was manipulated by giving different versions of source information and text presentation: In the credible condition, source information was varied by including banners of two well-respected university hospitals as well as of the German Federal Ministry of Health and the German Federal Ministry of Education and Research (see Eysenbach & Köhler, 2002; Rains & Karminkel, 2009). In contrast, in the non-credible conditions, these banners were replaced by (fictional) advertisements. In addition, the credible conditions provided source information in terms of the fictional author's name, place of work, and email contact information (see Eysenbach & Köhler, 2002); all fictional authors had an M.D. title (see Rains & Karminkel, 2009; Winter & Krämer, 2012). In the non-credible conditions, only a pseudonym (such as "Bea65" or "DJAlex71") was given as the author's name. With respect to text presentation, the credible website versions contained correct spelling. To give the impression of sloppiness, spelling errors were induced into the non-credible texts (see Fogg et al., 2002). We used typical typos that have no large impact on the comprehensibility of the text, such as incorrect capitalization, switching two letters, the omission of letters, the repetition of letters or syllables, and the substitution of letters by other letters that would be pronounced in the same way (see Kreiner et al., 2002).

All texts used in these manipulations had similar length and a similar amount of errors for the non-credible conditions. In contrast to study 1 to 4 we did not present a fully functional website. Instead, we used static screenshots in order to control exactly what webpages were visited and evaluated by the participants (screenshots can be requested via the corresponding author, one example is given in Appendix C.3).

The final version of the Web-CLIC as identified in study 1 and 2 (see Figure 3) was used as measuring instrument.

Procedure

Participants received an invitation via e-mail sent by the online panel PsyWeb. They were informed about involved researchers, anonymity, voluntariness and duration of the study. Participants were told that the aim of the research was to test the validity of the Web-CLIC, but were not given specific details. After providing some demographical data (e.g., age, gender, education level, Internet experience), they were instructed, that five screenshots of a randomly selected website would be shown to them and that they would have to answer five questions regarding it (i.e., the task was searching for information). Subsequently, participants were assigned to one out of eight possible conditions and first saw a screenshot of the website starting page, followed by four different subpages presented in a randomized order; on each screenshot, a question with regard to contents was presented. After that, the Web-CLIC and three additional measures (regarding usability, aesthetics, and recommendation, all not pertinent to the current study) were given in randomized order along with the instruction to refer items to all five webpages presented. Afterwards, three manipulation checks were performed, one for each manipulation (clarity, informativeness, and credibility). In each of those manipulation checks, two screenshots were presented randomly. While one screenshot in each case represented the condition that contained a high clarity, informativeness, and credibility, the other screenshot was drawn from an experimental condition in which one of those three constructs had been deliberately worsened. Participants had to indicate, which version they found more intelligible, of higher informativeness, or more credible. Finally, participants could request feedback about their performance and exclude their data from the subsequent analyses. They received disclosing information regarding the study, including that the presented website was fictional and might have been manipulated regarding its clarity, informativeness, and credibility. The participants were thanked and given the option of commenting on the study as well as to obtain a summary of the results. Field time of the study was from 07/09/2013 till 07/19/2013; completing it took about 20 minutes.

6.2. Results and discussion

Manipulation check

First, we checked whether the manipulations had worked. Participants were asked to indicate which of two presented pages of the website appeared more intelligible, informative or credible (see procedure). Thus, the actual percentages of answers that conformed to the manipulations were tested against a probability of .50 using an exact binomial test. Results showed that participants correctly identified 86% (for clarity; $p < .001$), 89% (for informativeness; $p < .001$) and 94% (for credibility; $p < .001$) of the manipulated websites. Thus, we assumed that manipulations worked quite well and as intended.

Experimental validation

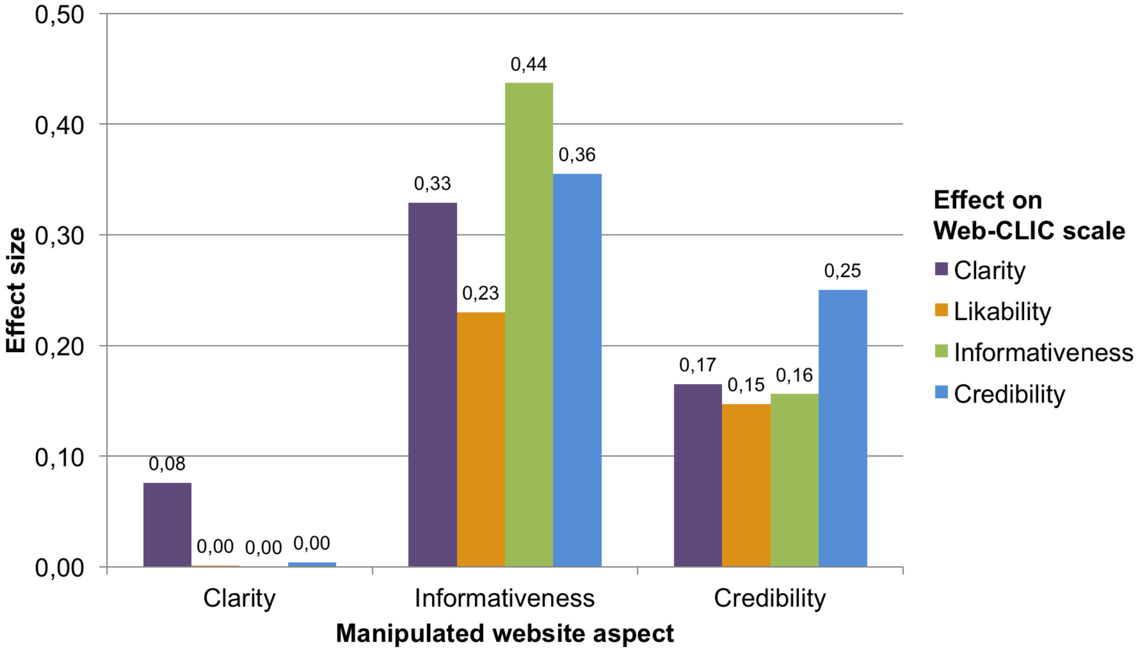
A MANOVA was performed to examine whether the Web-CLIC scores differ significantly as a consequence of the conducted manipulations. Thus, the three independent variables used were high vs. low clarity, high vs. low informativeness, and high vs. low credibility. The dependent variables were the Web-CLIC's four scales *clarity*, *likeability*, *informativeness,* and *credibility.* The model contained main-effects for the variables and their interactions. In order to describe the effects of the manipulation in more detail, four separate follow-up ANOVAs were calculated, each using one of the subscales as dependent variable.

As can be seen in Figure 8, all of the three different manipulations have significant multivariate main-effects and interactions on the Web-CLIC. The manipulations of credibility and informativeness revealed large effects, while the manipulation of clarity only led to a medium-effect (following the classification by Cohen, 1988). Multivariate interactions were significant but about an order of magnitude smaller than the main-effect. Importantly, they do not affect the interpretation of main-effects. The follow-up ANOVAs confirm the significant main-effects (see Figure 9). Specifically, the largest effect sizes were found for the informativeness manipulation, followed by credibility and clarity. Importantly we found that within each manipulation the strongest effects were always on the intended scales, i.e. the informativeness manipulation had the strongest effect on the informativeness scale, the credibility manipulation had the strongest effect on the credibility scale, and the clarity manipulation had the strongest effect on the clarity scale. In consequence, our findings confirm the idea, that the facets enclosed in the Web-CLIC jointly reflect subjective content perceptions, while each scale also carries a unique meaning. Especially, our manipulation of clarity aspects only affects ratings on the clarity scale. The performed manipulations, with respect to aspects of informativeness and credibility, to some extent affected all scales, but mostly the informativeness, as respectively the credibility scale. As we simultaneously manipulated source and message credibility, future research with a focus on message credibility only might lead to a clearer result pattern concerning this facet. In sum, the selective response of Web-CLIC scales to content features provides further evidence for construct validity.

Figure 8. MANOVA for the manipulations targeted at clarity, informativeness, and credibility (N = 567).

Manipulation	*F*	*df*	*p*	*partial* η^2
Clarity	28.245	4, 556	< .001	.171
Informativeness	115.269	4, 556	< .001	.453
Credibility	50.857	4, 556	< .001	.269
Clarity X Informativeness	2.91	4, 556	.02	.021
Clarity X Credibility	3.26	4, 556	.01	.023
Informativeness X Credibility	2.68	4, 556	.03	.018
Clarity X Informativeness X Credibility	2.90	4, 556	.02	.020

Figure 9. Effect sizes (eta-squared) for the three different experimental manipulations (main effects of high vs. low condition).



7. STUDY 6. FURTHER VALIDATION AND USEFULNESS OF THE WEB-CLIC

The main goals of study 6 are to perform additional validations and to demonstrate the usefulness of the Web-CLIC. With regard to the first aim, we tested if the Web-CLIC is influenced by the task formats given in the prior studies (free exploring was used in study 1, 2, and 4; search tasks were used in study 3 and 5). Furthermore, we provide evidence for convergent validity of the Web-CLIC credibility scale, by comparing it to an established credibility measure (as this was missing in study 4). With regard to the second aim, we tested whether Web-CLIC ratings are related to intentions and actual behavior: the donation of money to one of three different organizations (i.e., predictive validity). Specifically, we tested whether the Web-CLIC explained variance above and beyond global ratings of websites.

7.1. Method

Participants

A total of 268 participants took part in this web-based study; 147 were female (54.9 %), 120 male (44.8 %). Ages ranged from 14 to 77 years (*M* = 47.68, *SD* = 13.34). The education level of 66.8 % of the participants was Abitur (German university entrance qualification) or higher. On average, the participants had been using the Internet for 17.24 years (*Min* = 5, *Max* = 30, *SD* = 4.56) and stated an active use of on average 2.48 hours a day (*Min* = 0.15, *Max* = 15, *SD* = 2.00). Participants took part voluntarily and on an anonymous basis; they received no compensation but could request a summary of the study's results.

Stimulus materials and measures

We used a 3 (Stimuli) x 2 (Tasks) mixed within-between design. All subjects rated the same three websites of nonprofit organizations in random order. Using search engines, we selected typical organizations supporting education and access to knowledge. Tested websites were of the initiatives "Studenteninitiative Weitblick e.V." (https://weitblicker.org/), "Suma e.V. – Verein für freien Wissenszugang" (http://suma-ev.de/), and "VFoB -Verein zur Förderung der offenen Bildung e.V." (http://vfob.org/; screenshots can be requested via the corresponding author). Between subjects, task format was manipulated: One group (n = 137) was instructed to freely explore the given websites, the second group (n = 131) was asked to search for information and answer three questions about each website:

1. What is the aim of the organization – which people ought to be supported? [Anchored with "pupils", "students", "participants of specific projects", "all people", "none of this is correct"]
2. Where can detailed information on the aim of the organization be found? [Copy in the URL]
3. Who is the chairperson of the organization? [Copy in the name]

As measuring instrument, the final version of the Web-CLIC (see Figure 3) was used. In addition, participants were asked to rate the websites on the credibility scale of Appelman and Sundar (2016). This scale consists of three items, showed good reliability (α = .87) as well as content, criterion and construct validity (see Appelman & Sundar, 2016, p. 72). The overall website score was assessed with the same six-point grading scale as used in study 4. Two additional global items were given ("The website is of high quality." and "I like the website"), using the same Likert scale format as the Web-CLIC.

Procedure

Participants were invited via e-mail through the online panel PsyWeb; participants of prior Web-CLIC studies were excluded automatically. The study was announced as general website evaluation study. On the first two survey pages, all participants were informed about involved researchers, anonymity, voluntariness and duration of the study. Participants were included if none of the three organizations was familiar to them. After providing demographical data as in prior studies, the three websites were randomly presented to the participants, along with a task (exploring versus searching), and the request to answer the given measures with respect to each website. After that, participants received a forced choice item to which of the three organizations an amount of €100 should be donated (money was provided by the investigators). Additionally, they were asked which amount of money they potentially would donate themselves to each organization (€0 was a possible answer). Finally, participants received further information on the study and had the opportunity to exclude their data from subsequent analyses. They were thanked and given the option of commenting on the study as well as to obtain a summary of the results. The study was available online from 01/09/2017 till 01/26/2017; completing it on average took about 20 minutes.

7.2. Results and discussion

First, we tested whether the task affected the Web-CLIC total or subscale-scores. Since all participants rated all websites, we used a multilevel model to account for repeated measures. Specifically, we calculated five separate linear mixed effect models to predict the sum score and four subscale-scores using site and task as fixed effects and participant as random effect. Of these five models, only the model for the clarity subscale showed a significant effect for task, i.e. participants who worked on the search task gave significantly higher clarity ratings (*M* = 4.77, *SD* = 1.30) than participants in the free exploration condition (*M* = 4.53; *SD* = 1.48). Because this represented only a small effect (*d* = 0.17), and was only observed for one of the three websites, we treat this as a random result rather than a systematic trend. For a detailed investigation of this issue, readers are referred to Dames and colleagues (under review). In here, a study using the Web-CLIC found that the task (free browsing vs. goal-directed searching) had an effect on the strength of the influence of content perception on intentions to recommend or revisit, but not on the overall impression of a website or the directions of the effects found.

Second, we tested the correlations between the Web-CLIC credibility scale and the credibility scale of Appelman and Sundar (2016) for the three websites and the two task-conditions. As can be seen in Figure 10, all six correlations can be considered large and highly significant, providing further evidence of convergent validity.

Figure 10. Within-participant correlations between the Web-CLIC scale credibility and the credibility scale by Appelman and Sundar (2016) for the three websites and the two tasks separately.

		Task	
		Free exploration (n = 137)	Search for information (n = 131)
Organization	Weitblick e.V.	.804	.864
	Suma e.V.	.853	.876
	VFoB e.V.	.830	.815

Note: All correlations are significant with *p* < .001.

Third, we wanted to establish that the website that received the highest global Web-CLIC rating was also the one that participants voted to donate money to (Figure 11). For this, we combined data from both task-conditions and determined the website that received the highest Web-CLIC rating for each participant. For nine participants, the highest Web-CLIC score was tied, i.e. two websites got a similarly high rating. In these cases, we randomly chose which of the sites got the highest rating (we repeated this procedure to ensure that it did not affect the results). In order to show the association between this rating and the forced-choice between one of the organizations, a chi-square test was used. This indicated that there was a significant association between content ratings and the decision to which organization money should be donated (χ^2 (4) = 118.03; p < .001), showing a "large effect" (*Cramers V* = .47) according to Cohen's guidelines (Cohen, 1988). Repeating this analysis 1000 times yielded significant and "large" effects in all repetitions, demonstrating predictive validity of the Web-CLIC.

Figure 11. Relationship between highest content ratings and the decision to donate money for a specific organization (N = 268).

		Donation recipient		
		Weitblick e.V.	Suma e.V.	VFoB e.V.
Highest content rating	Weitblick e.V.	110	20	23
	Suma e.V.	18	60	4
	VFoB e.V.	8	9	16

Fourth, we tested whether the Web-CLIC explains variance above and beyond a simple global item in the two task groups. For this, we used a logistic regression model to predict whether or not a participant would donate money for a specific organization. Since all participants rated all websites and indicated how much money they wanted to give to each organization, we used a multilevel model to account for the fact that each participant contributed three observations to the dataset. Our critical comparison involved two models. The first used the overall grade only to predict whether or not a participant intended to donate money to this organization. The second model used the overall grade and the Web-CLIC scales to predict the intention to donate money to this organization. The two models were compared using likelihood ratio-tests, and variance explained was measured using Tjur's D (Tjur, 2009). We found that the second model (*Tjur's D* = .74) showed a much better fit to the data than the first model (*Tjur's D* = .46; χ^2 (4) = 154.43, p < .001). Similar results were found for the two alternative global items: The Web-CLIC scales predicted user's intentions to donate above and beyond these items.

In sum, we demonstrated with this study the high usefulness of evaluations gathered with the Web-CLIC in predicting not only intentions but as well actual user behavior.

8. STUDY 7. BENCHMARKS AND OPTIMAL CUT POINTS FOR THE WEB-CLIC

In practical use, it will be helpful to consider precise Web-CLIC values for specific comparisons with a tested website, thus study 7 aims at providing benchmarks. In addition, we calculated optimal cut points as an orientation if a website should be assessed on a general level or for situations when no benchmark is available (see Hirschfeld & Thielsch, 2015). Furthermore, benchmarks do not offer information on the relevance of specific cut points, for example even if the content of a specific website receives above-average ratings, that does not imply that users are satisfied with the presented content on the website. For the cut point analyses, we combined data from nine different website evaluation studies: the data from study 2, study 3 (only T1) and study 4 of the current paper as well as data from six additional, currently unpublished, studies from our research group. In these studies, the Web-CLIC was applied together with an overall website evaluation, that we used as criterion for the cut point analyses. For the benchmark analysis, we included additional data from study 6 of the current paper, as well as data from Dames and colleagues (under review), Thielsch and Thielsch (under review), Thielsch and Wirth (2017), and one additional, currently unpublished, study from our research group.

8.1. Method

Participants

A combined sample of 5363 participants was used for benchmark analysis, among them 2863 females (53.4 %) and 2500 males (46.6 %). Of those, data of 3545 participants could be used in cut point analyses (55.8 % females, 44.2 % male). Ages ranged from 14 to 89 years (*M* = 34.49, *SD* = 13.89 respectively *M* = 33.43, *SD* = 14.35 in cut point analyses). The education level of 59.3 % of the participants was Abitur (German university entrance qualification) or higher (67.1% in cut point analyses); for 17.2 % (respectively 11.6 % in cut point analyses) specific data for the educational level were not available. On average, the participants had been using the Internet for 12.70 years (*Min* = 1, *Max* = 35, *SD* = 4.75; data available for n = 4833) and stated an active use of on average 2.62 hours a day (*Min* = 0.01, *Max* = 16, *SD* = 2.01; data available for n = 5099). Participants included in the cut point analyses had been using the Internet for 11.97 years (*Min* = 1, *Max* = 30, *SD* = 4.52; data available for n = 3539) and stated an active use of on average 2.49 hours a day (*Min* = 0.01, *Max* = 16, *SD* = 1.87; data available for n = 3449). In all studies, participants took part voluntarily and on an anonymous basis. That is why we cannot rule out that some of the participants might have took part twice (yet, additional cut point analyses with the largest unique sample of 1226 participants resulted in very similar results compared to the whole sample). Mostly they received no compensation but could request a summary of the study's results, in some studies they could take part in a lottery of vouchers or students could receive course credits for participation.

Stimulus material and procedure

In each study, participants were informed about its objective, involved researchers, anonymity, voluntariness and duration. After providing demographic information, participants were usually randomly assigned to one or two fully functional websites from the respective stimulus set; only in one study participants were asked to evaluate more than three websites. In sum, 7379 ratings on 120 websites and additional eight online annual business reports (see Thielsch & Wirth, 2017) were analyzed (respectively in cut point analyses: 4246 ratings on 100 websites). Each website belonged to one of ten different categories (see Appendix C.2), and on average was evaluated by 58.13 participants (*Min* = 13, *Max* = 481, respectively *M* = 42.46 participants with *Min* = 13 and *Max* = 204 in cut point analyses). Mostly, the website in question was presented within a split screen, the Web-CLIC items were presented in a smaller upper panel. In studies that were included in cut point analyses, an additional six-point grading scale (1 = "very good", 2 = "good", 3 = "satisfactory", 4 = "adequate", 5 = "poor", 6 = "unsatisfactory") was applied. At the end of each study, participants could exclude their data from the subsequent analysis and were thanked.

8.2. Results and discussion

Influences of age, gender, and education level

Before calculating benchmarks, we first checked the extent to that the Web-CLIC is influenced by age, gender or education. Correlation between age, education level, and the Web-CLIC scores were very small ($r \leq -.075$), but partly significant due to sample size (see Appendix D.1). Yet, even the biggest variance explained by one of these correlations is far below 1 % (exactly 0.563 % for education level with likeability).

Furthermore, the Web-CLIC in general proved to be robust towards gender effects: There is only a small difference of 0.051 between men and women in the Web-CLIC sum score ($M_{\text{Women}} = 4.448$; $M_{\text{Men}} = 4.397$). A standardized mean difference effect size of $d = 0.043$ indicates, that this gender effect has practically little to no relevance. The same accounts for all four Web-CLIC subscales:

- Clarity: $M_{\text{Women}} = 4.866$; $M_{\text{Men}} = 4.760$, $d = 0.083$
- Likeability: $M_{\text{Women}} = 3.620$; $M_{\text{Men}} = 3.529$, $d = 0.058$
- Informativeness: $M_{\text{Women}} = 4.572$; $M_{\text{Men}} = 4.518$, $d = 0.039$
- Credibility: $M_{\text{Women}} = 4.738$; $M_{\text{Men}} = 4.784$, $d = -0.043$

Thus, in general, website evaluation effects of age, gender or education could be neglected. Still, in specific situations it might be important to keep an eye on such variables: For example, when analyzing special target groups or evaluations of specific web contents with relation to age, gender or education.

Benchmarks for different website categories

Clear differences appear in a MANOVA with website category as independent variable, Web-CLIC scores as dependent variables and age, gender and education level as covariates: $F_{website\ category}$ (40, 19136) = 39.013, $p < .01$, $\eta^2 = .075$. Thus, we calculated Web-CLICs means and standard deviations separately for each website category. This benchmark (Figure 12) can be used to compare results from a newly tested website with the respective category. Yet, one has to keep in mind that in most studies participants were randomly assigned to websites that have been unknown to them. Thus, the benchmark reflects the evaluation of random web user, not of people highly familiar with a given website (such as registered costumers of an e-commerce website). In addition, in some categories, only few (less than five) websites were tested and thus results should be considered as preliminary. In such cases, or if no category in the benchmark is fitting at all, we recommend using the general cut points presented in the following section.

Figure 12: Benchmark of the Web-CLIC: Overall means as well as means for each scale as functions of a website category

Category	WEB-CLIC sum score		Clarity		Likeability		Informativeness		Credibility	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Download & Software (m = 4; n = 107)	3.55	1.10	4.03	1.34	2.62	1.45	3.65	1.29	3.90	1.18
E-Commerce (m = 12; n = 934)	4.79	1.07	5.04	1.15	4.03	1.49	4. 91	1.19	5.17	1.17
Entertainment (m = 4; n = 184)	2.89	0.95	4.46	1.43	2.00	1.21	2.15	1.25	2.96	1.42
E-Learning (m = 5; n = 90)	4.62	0.99	4.74	1.35	3.58	1.49	4.82	1.28	5.32	0.97
E-Recruiting & E-Assessment (m = 26; n = 1617)	4.62	1.14	4.95	1.24	3.84	1.50	4.79	1.25	4.93	1.32
Information (m = 14; n = 1437)	4.59	1.23	4.88	1.27	3.77	1.61	4.82	1.36	4.89	1.41
Presentation & Self-portrayal: Websites (m = 40; n = 2361)	4.41	1.17	4.70	1.31	3.57	1.61	4.47	1.31	4.89	1.25
Presentation & Self-portrayal: Online business reports (m = 8; n = 165)	4.41	1.03	4.80	1.13	3.53	1.45	4.66	1.16	4.64	1.08
Search engines (m = 4; n = 125)	4.23	0.97	4.87	1.19	3.39	1.35	4.31	1.18	4.35	1.11

Web portals (m = 5; n = 226)	3.78	1.07	4.03	1.21	2.65	1.32	4.13	1.26	4.30	1.35
Weblogs and Social Sharing (m = 6; n = 133)	3.64	1.25	4.38	1.36	2.95	1.58	3.41	1.45	3.82	1.34
Sum score (m = 128; n = 7379)	4.45	1.20	4.79	1.28	3.63	1.59	4.57	1.37	4.82	1.35

Note: *M* = mean, *SD* = standard deviation, m = number of evaluated websites in one category, n = number of participants. Evaluations of online annual business reports were included as a subcategory of the presentation and a self-portrayal category, representing a special form of typical web-based corporate communications (see Thielsch & Wirth, 2017).

Cut point analyses

In order to establish meaningful cut points for the interpretation of the Web-CLIC, we used receiver-operating-characteristic (ROC) based methods (see Hirschfeld & Thielsch, 2015). These methods identify those cut points for the content-ratings, that differentiate best between websites that were overall rated as good (grades 1 or 2) and websites that were overall rated as not good (grades 3, 4, 5, or 6). Specifically, these methods entail calculating the sensitivity and specificity for all possible cut points on the sum score and the subscales. Sensitivity refers to the percentage of good websites that actually get a scale score larger than the cut point. Specificity refers to the percentage of bad websites that actually get a scale score smaller than the cut point. The cut point that yields the highest sum of sensitivity + specificity (i.e. Youden-index) is identified as optimal. We found that websites that were overall rated as good received a higher Web-CLIC rating (*M* = 5.13) than websites rated as not good (*M* = 3.80; *t* (4244) = -44,08, *p* < .001, *d* = 1.55). Furthermore, the Web-CLIC showed an area under curve (AUC) of .848 (95% CI: .836 - .860) indicating a good classification of the websites based on the overall rating. The cut point that was defined as optimal was 4.58, i.e. content ratings below 4.58 indicate a "bad" website, while content ratings higher than 4.58 indicate "good" websites (see Figure 13 and Appendix D.2). Using this cut point to determine if a website is good or bad would result in 77 percent of the good websites identified as good (sensitivity) and 79 percent of the bad websites identified as bad (specificity). Testing the variability of the optimal cut points using bootstrapping showed that this cut point was selected as optimal in 57.47% of the pseudo-samples. Other cut points that were selected as optimal were 4.5 and 4.67 (selected in 23.02%, respectively, 18.78% of the pseudo-samples). This indicates that we were able to estimate the optimal cut points with a relatively high precision.

Results concerning the Web-CLIC subscales were very similar to the results for the sum score (see Figure 13). Specifically, the individual subscales also showed large differences between good and bad websites (Cohens *d* between .98 and 1.4) and a good classification (AUC between .74 and .81). Furthermore, the optimal cut points for the subscales also showed acceptable levels of sensitivity and specificity, with only the credibility scale showing low specificity. Cut points for the subscales showed a little more variety than the sum scale with the highest cut point (5.33) for the clarity scale and the lowest one (4.00) for the likeability scale.

Overall, the results indicate that a binary interpretation of the Web-CLIC based on the presented cut points is feasible at the level of the sum score, as well as with regard to the individual subscales. Yet, the AUC was only acceptable, maybe due to the limited reliability of the overall rating that was assessed with a single item. Further research using alternative gold-standards (see Hirschfeld & Thielsch, 2015) is needed to test the generalizability of this cut point. The high agreement between bootstrapping samples indicates some stability of this cut point based on the fairly large sample size. As a consequence, aiming for an overall Web-CLIC rating of 4.58 or higher would be a recommendable goal for most practical applications. If a specific aspect (e.g., clarity) is targeted, we recommend interpreting the findings using the respective cut point given in Figure 13.

Figure 13. Optimal cut points for the Web-CLIC, including information about effect size for differences between websites classified as good and bad, AUC, sensitivity, and specificity.

Scale	Effect-size (Cohen's *d*)	AUC	Optimal cut point	Sensitivity	Specificity
Web-CLIC sum score	1.55	.85	4.58	.77	.79
Clarity	1.35	.81	5.33	.70	.79
Likeability	1.40	.81	4.00	.71	.78
Informativeness	1.22	.79	5.00	.69	.74
Credibility	.98	.74	4.67	.79	.56

9. GENERAL DISCUSSION

The aim of the present research was to develop a sound measure assessing subjective web content perceptions. In a series of seven studies we have demonstrated reliability, as well as various aspects of validity of the resulting Web-CLIC questionnaire, and provided guidelines for practical application. In the following we (1) describe the individual facets assessed with the Web-CLIC, (2) discuss the quality of the measure, (3) develop practical implications, (4) highlight limitations, and (5) sketch avenues for future research.

9.1. Facets of users' subjective perceptions of website content

Web-CLIC items were based on the existing literature with a focus on facets of website content that average website users can comment on. We followed an interactionist view on content perceptions, thus Web-CLIC facets refer to idiosyncratic evaluations of web content objects. In contrast to algorithmic measures of website content (e.g., word counts or syntactical analysis), the Web-CLIC focuses on subjective perceptions. The final Web-CLIC consists of four scales that measure different subjective content facets, jointly representing a general factor of subjective perception of website content.

The facet *clarity* relates to the extent information is presented on a website in a clear, comprehensible and easy to understand manner. Thus, clarity is sometimes labelled as "comprehensibility" (e.g., Elling et al., 2007), "ease of understanding" (e.g., Delone & McLean, 2003), "intelligibility" (e.g., Thielsch, 2008) or "understandability" (e.g., Caro et al., 2008). Importance of this facet is already stressed in common models (e.g., Delone and McLean, 2003; Dinet et al., 2012), and it is consequently enclosed in several other measures of website content (e.g., Aladwani & Palvia, 2002; De Marsico & Levialdi, 2004; Thielsch, 2008). The Web-CLIC clarity scale comprises how users evaluate the intelligibility of web contents, the extent to which these are presented in a clear and concise manner, and the comprehensibility of the language used.

The facet *likeability* grasps users' interests in a website, and his or her emotional perceptions of the content presented on a website. Likeability of web content is also discussed under the labels "perceived attractiveness" (e.g., Caro et al., 2008) or "entertainment" (e.g., Huizingh, 2000) and is enclosed in some prior measures (Kang & Kim, 2006; Thielsch, 2008). The Web-CLIC likeability scale assesses the amount of interest, excitement and joy website content can trigger in a user. Thus, a users' general emotional evaluation of the website content is indicated on this scale.

The facet *informativeness* refers to the perceived amount of useful and valuable information given on a website. This facet is most strongly related to the general factor of subjective website perception found in our studies, which indicates its central role. This result is in line with the frequent use of informativeness scales and items in prior measures (e.g., Chakraborty et al., 2005; Hausman & Siepke, 2009; Kang & Kim, 2006; Lin, 2007; Rahimnia & Hassanzadeh, 2013; Shukla et al., 2010). The Web-CLIC informativeness scale comprised items related to the quality, usefulness, and value of the information presented on a website.

The facet *credibility* is a global scale assessing the believability of information presented on a website. Due to its inherent importance for a broad range of website operators, credibility is often researched and part of many measures of website perceptions (e.g., De Wulf et al., 2006; Flanagin & Metzger, 2000; Fogg et al., 2001; Hong, 2006; Metzger, 2007; Wathen & Burkell, 2002). The Web-CLIC credibility scale refers to general aspects of authenticity, reliability, and trustworthiness of a given website content.[5]

9.2. Objectivity, reliability, and validity of the Web-CLIC

The evaluation of psychometric criteria focused on reliability and validity. However, with the Web-CLIC questionnaire being a standardized measure, objectivity in the test situation can easily be achieved, especially when it is carried out in a computer-based manner. Moreover, since objectivity is a necessary condition for reliability, positive evaluations in terms of reliability also indicate a high objectivity. In fact, high values for internal consistency are found, clearly exceeding those of prior measures such as the ICTQ (Ozok & Salvendy, 2001) or the WWI (Thielsch, 2008). This is notable, in particular when considering the brevity of the Web-CLIC. Furthermore, little is known about the stability of web content perceptions over several days or weeks and no such data was available for prior instruments. Nevertheless, the Web-CLIC sum score and several sub-scales performed well in respective analyses (see study 3), showing sufficient to good retest values. In sum, we can state a high reliability of the Web-CLIC measure.

Furthermore, we found evidence for a high validity of the Web-CLIC by demonstrating high correlations to convergent and concurrent criteria. Correlations to divergent criteria were lower, however, sometimes still higher than expected. In consequence, we performed an experimental validation which shows the sensitivity of the clarity, informativeness, and credibility scales for corresponding changes in website content. This provided evidence for construct validity that is highly relevant to practitioners who want to use the Web-CLIC to assess the impact of design alterations on perceptions. Moreover, the Web-CLIC is able to differentiate between different websites (as shown in study 4) and was not influenced by basic user demographics (as shown in study 7). Finally, practical utility of the Web-CLIC is demonstrated not only by its high correlations to concurrent criteria (see study 4), but also by its capability in predicting user intentions and actual user behavior (see study 6). At the same time, the Web-CLIC offers some advantages over single-items measures of content quality or overall quality: First, as the experimental validation has shown, changing specific aspects of websites may affect some facets of content perceptions but not others. Compared to a single item the Web-CLIC gives more detailed information on what aspects of a website are affected (respectively need to be improved). Second, as we demonstrated in study 6, the Web-CLIC has incremental validity above and beyond single-item measures as it improves the prediction of intended behavior.

[5] If sub-facets of credibility are of interest, we recommend the use of a more specific measure further differentiating this facet (e.g., Chung, Nam, & Stefanone, 2012).

9.3. Interpretation of the Web-CLIC and practical implications

The Web-CLIC is a short measure. After exploring a website, most people need less than two minutes to answer the twelve items. Additionally, the items are easy to understand, no specific knowledge or expertise and almost no instruction is needed (as instruction, in our studies we just asked participants to rate a given website). We presented the 12 items with a Likert scale ranging from 1 (totally disagree) to 7 (totally agree), all anchor points were verbally labeled (see supplements). The Web-CLIC was validated on samples of adults and adolescents older than 14 years and thus could be applied to those age groups. So far, we have no experiences with respect to the application of the Web-CLIC in studies with children. In practical use, we recommend testing a fully functioning version of the website in question and the use of relevant tasks (e.g., searching or browsing tasks) to simulate typical use. Usually, items should not be changed in wording, except for minor adjustments to ensure comprehensibility and perfect fit to the target stimulus. For example, Thielsch and Wirth (2017) analyzed web-based annual reports with the Web-CLIC and changed the term “website” to “report” in eight items, still the questionnaire was well applicable and showed good reliability. However, items should not be completely removed, as the Web-CLIC scales are already very short and further reductions can compromise psychometric quality. If a specific facet should be focused solely, it is possible to use the respective single scale of the Web-CLIC alone, as all four subscales showed high reliability and validity.

When the user survey is done, the analysis of answers given on the Web-CLIC starts with overall mean, as well as means for each subscale. These can be calculated in a way, that high scores represent a high value on the respective scale. In order to calculate the means of each scale, the single values of each subscale are added up, and the resulting sum is divided by three (i.e. the number of items for the subscales). The general factor, the overall mean of the questionnaire, can be calculated by adding all scale values and dividing them by four – or by dividing the sum of all items by 12. We recommend interpreting the Web-CLIC on the level of the four facets and the sum score only, but not on single item level.

When interpreting Web-CLIC mean values, it is essential to consider the subjective character of the evaluations. For example, a high value on the scale *informativeness* does not indicate a particularly well-texted and informative website, but a positive evaluation of the perceived informativeness by the website users. This way of interpretation should be applied analogously for the other scales. Regarding the interpretation of the overall mean, a low value indicates a negative evaluation of the website’s content in general. Furthermore, the Web-CLIC presented itself as generally robust against bias effects caused by age, gender, or educational level. For practical use, we determined optimal cut points for the Web-CLIC, indicating that sum score values above 4.58 are desirable (for respective values for the subscales see Figure 13). Additional benchmark values for ten different content domains of websites (see Figure 12) further assist in the practical interpretation of evaluations performed with the Web-CLIC. If applicable, we recommend the use of the Web-CLIC in direct comparisons, for example, between prior and novel website versions, with other topic-related existing websites, or different prototypes. In practice, aiming at higher values compared to competing websites might be easier than trying to reach the top of each Web-CLIC scale.

9.4. Limitations and future research

There are several limitations one has to keep in mind when interpreting the present findings, some of which highlight possible avenues for future research. First, as mentioned above, the Web-CLIC is limited to the evaluation of distinct subjective content facets. Thus, it is desirable that future research further investigates the connection and possible overlaps between the many different facets discussed in research (see Appendix A), as well as the interplay with related constructs such as usability and aesthetics (e.g., Cober at al., 2003; Thielsch et al., 2014). This would enable a better understanding of underlying cognitive processes in website perception. In practical use, it might be very interesting to combine subjective measures of web content with results from automatic algorithms. While some of the Web-CLIC scales already imply starting points for website improvements, practitioners will further profit from such findings, showing the consequences of content improvements on a user level.

Second, more studies are needed that relate perceptions of websites to actual behavior. We found that content evaluations predicted decisions to donate money to a charity, but it is unknown if perceptions of content are similarly related to user behaviors in other relevant domains such as e-commerce, e-health or e-learning. For example, one important aspect of web-based health interventions is dropout (von Brachel et al., 2014). One could test whether perceptions of content predict whether or not participants complete a treatment. While this would show the general significance for individuals, it would be at least as important to show that Web-CLIC facets are a relevant predictor across different interventions. For example, showing that interventions which are on average rated as more credible are more effective, would provide a strong rationale for designers of interventions to improve on this aspect. This could be done by either systematically manipulating aspects of health interventions or in the form of a meta-analysis across several interventions provided that these use similar measures for subjective perceptions of content. We hope that the Web-CLIC will be routinely used to assess content enabling such comparisons across studies.

Third, the construction of the Web-CLIC included more than 3,100 participants evaluating 60 websites from a broad variety of domains. But still, neither the tested websites nor the participants can be seen as perfectly representative for the enormous number of existing websites and web users. Thus, replications of our studies and further investigations of validity and applicability of our measure are highly welcome.

Fourth, we would like to highlight that all tested participants shared a common cultural background. In the construction of the measure, we used a German version that afterwards was systematically translated into English by a native speaker and successfully applied in the study of Dames and colleagues (under review). Culture is a possible cause of bias, as it plays an important role in website content (see Fletcher, 2006), and cultural differences are even found on the level of content features (e.g., Robbins & Stylianou, 2003; Zhao et al., 2003). Thus, future studies should investigate cultural effects of subjective content perceptions, as well as possible effects on the Web-CLIC. Different language versions of the Web-CLIC measure are very welcome as well. In doing so, it would be important to test whether cultural differences are due to how the measure operates in different cultural contexts (i.e. lack of measurement invariance) or real differences in how the same aspects of websites are perceived in different countries.

Fifth, in all studies except for study 5, fully-functional websites were used as stimulus material. The use of fully-functional stimuli increases realism of test situations at the cost of experimental control. In contrast, the use of non-interactive screenshots can lead to superficial processing and halo-effects. Yet, to the best of our knowledge, empirical evidence for this issue was found for usability assessments only (see Thielsch et al., 2015), but not for content. In addition, when it is vital that all participants see and read the exact same information, screenshots are still the best way to conduct an experimental design. In such scenarios, we suggest that researchers use methods such as specific tasks to avoid superficial processing (as done in study 5). In general, fully-functional websites are of great value in adding external validity to evaluation studies, but in an experimental investigation of content perception, screenshots are a good way to ensure profound processing of information.

Sixth, we have not investigated how subjective web content evaluations develop over time and what aspects of websites affect the possible changes. It may be relatively easy to find aspects that determine how first-time users perceive a website (see Tuch et al., 2012), however it may be much harder to change perceptions of returning users. Further research on the interplay between web content perceptions and other perceptions of websites focusing on the timeline of use could be promising (see Thielsch et al., 2014). In addition, such research could include systematic variations of web design features to investigate causal relationships.

Seventh, the answer time for the Web-CLIC measure is short, but there might be situations where a very brief measure is needed, for example when conducting a screening or a manipulation check. Future research should aim at the creation of such a short form of the Web-CLIC.

9.5. Conclusion

The present research focused on subjective perceptions of web content and the measurement of them with the newly developed Web-CLIC. This measure comprises four scales – *clarity*, *likeability*, *informativeness*, and *credibility* – jointly representing a general factor, the *subjective perception of content.* In extensive quality tests, the Web-CLIC showed high reliability and construct validity. Particularly, as shown in an experimental validation, Web-CLIC scales are sensitive to corresponding changes in website content. Furthermore, the Web-CLIC is capable of predicting user intentions and behavior. Consequently, we highly recommend the use of the measure in future research and provided additional interpretation aids such as optimal cut points and benchmarks to facilitate its application in practice. In sum, the Web-CLIC is a sound measure of high value, allowing for a precise evaluation of users' subjective content perceptions.

Acknowledgments. Special thanks to Julian Breuer, who put a lot of effort in conducting study 1 and 4, as well as to Patrick Boertz for his great work done in study 5. Additionally, we would like to thank Dustin Witte for his assistance in conducting study 3, Jan-Hendrik Spieth for his help in the translation of Web-CLIC items, Friederike Jansen for her help in data preparation and Stefan Schmukle for his support and valuable suggestions. Finally, we thank Niki Kittur and four anonymous reviewers for their very helpful feedback provided on previous versions of this manuscript.

Authors' Mini-bios:

Meinald T. Thielsch (thielsch@uni-muenster.de, www.meinald.de) is a psychologist with an interest in human-computer interaction, user experience, and online research; he is an Akademischer Rat (tenured faculty member) at the Department of Psychology, University of Münster.

Gerrit Hirschfeld (g.hirschfeld@hs-osnabrueck.de, www.gerrithirschfeld.de) is a psychologist with an interest in statistics and research methods; he is professor for quantitative Methods in the Faculty of Business Management and Social Sciences, University of Applied Sciences Osnabrück.

REFERENCES

Abdinnour-Helm, S. F., Chaparro, B. S., & Farmer, S. M. (2005). Using the End-User Computing satisfaction (EUCS) instrument to measure satisfaction with a Web Site. *Decision Sciences*, *36*(2), 341–364. doi:10.1111/j.1540-5414.2005.00076.x

Agarwal, R., & Venkatesh, V. (2002). Assessing a Firm's Web Presence: A Heuristic Evaluation Procedure for the Measurement of Usability. *Information Systems Research, 13*(2), 168–186. doi:10.1287/isre.13.2.168.84

Ahn, T., Ryu, S., & Han, I. (2007). The impact of Web quality and playfulness on user acceptance of online retailing. *Information and Management, 44*(3), 263–275. doi:10.1016/j.im.2006.12.008

Aladwani, A. M. (2002). The development of two tools for measuring the easiness and usefulness of transactional Web sites. *European Journal of Information Systems, 11* (3), 223-234. doi:10.1057/palgrave.ejis.3000432

Aladwani, A. M., & Palvia, P. C. (2002). Developing and validating an instrument for measuring user-perceived web quality. *Information and Management, 39*(6), 467–476. doi:10.1016/S0378-7206(01)00113-6

Appelman, A., & Sundar, S. S. (2016). Measuring Message Credibility: Construction and Validation of an Exclusive Scale. *Journalism & Mass Communication Quarterly*, *93*(1), 59–79. doi:10.1177/1077699015606057

Aranyi, G., & van Schaik, P. (2016). Testing a model of user-experience with news websites. *Journal of the Association for Information Science and Technology*, *67*(7), 1555–1575. Doi:10.1002/asi.23462

Bao, S., Xue, G., Wu, X., Yu, Y., Fei, B., & Su, Z. (2007). Optimizing web search using social annotations. *Proceedings of the 16th International Conference on World Wide Web - WWW '07*, 501-510. doi:10.1145/1242572.1242640

Baumgartner, H., & Homburg, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International journal of Research in Marketing, 13*(2), 139-161.

Bölte, J., Hösker, T., Hirschfeld, G. & Thielsch, M. T. (2017). Electrophysiological correlates of aesthetic processing of webpages: A comparison of experts and laypersons. *PeerJ*, 5:e3440. doi:10.7717/peerj.3440.

Braddy, P. W., Meade, A. W., Michael, J. J., & Fleenor, J. W. (2009). Internet Recruiting: Effects of website content features on viewers' perceptions of organizational culture. *International Journal of Selection and Assessment*, 17(1), 19–34. doi:10.1111/j.1468-2389.2009.00448.x

Brin, S., Motwani, R., Page, L., & Winograd, T. (1998). What can you do with a web in your pocket?. *IEEE Data Engineering Bulletin*, *21*(2), 37-47.

Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N., & Hullender, G. (2005, August). Learning to rank using gradient descent. In *Proceedings of the 22nd international conference on Machine learning* (pp. 89-96). ACM.

Cai, D., Yu, S., Wen, J.-R., & Ma, W.-Y. (2003). Extracting content structure for web pages based on visual representation. *Proceedings of the 5th Asia-Pacific Web Conference on Web Technologies and Applications*, 406–417. doi:10.1007/3-540-36901-5_42

Cao, M., Zhang, Q., & Seydel, J. (2005). B2C e-commerce web site quality: an empirical examination. *Industrial Management & Data Systems, 105*(5), 645–661. doi:10.1108/02635570510600000

Caro, A., Calero, C., Caballero, I., & Piattini, M. (2008). A proposal for a set of attributes relevant for Web portal data quality. *Software Quality Journal,* 16 (4), 513-542. doi:10.1007/s11219-008-9046-7

Casaló, L. V., Flavián, C., & Guinalíu, M. (2007). The role of security, privacy, usability and reputation in the development of online banking. *Online Information Review, 31*(5), 583–603. doi:10.1108/14684520710832315

Chakraborty, G., Srivastava, P., & Warren, D. L. (2005). Understanding corporate B2B web sites' effectiveness from North American and European perspective. *Industrial Marketing Management, 34*(5), 420–429. doi:10.1016/j.indmarman.2004.09.008

Choi, S. M., & Rifon, N. J. (2002). Antecedents and consequences of Web advertising credibility: A study of consumer response to banner ads. *Journal of Interactive Advertising*, *3*(1), 12–24. doi:10.1080/15252019.2002.10722064

Chung, C. J., Nam, Y., & Stefanone, M. A. (2012). Exploring online news credibility: The relative influence of traditional and technological factors. *Journal of Computer-Mediated Communication*, *17*(2), 171–186. doi:10.1111/j.1083-6101.2011.01565.x

Clarke, C. L. A., Kolla, M., Cormack, G. V, Vechtomova, O., Ashkan, A., Büttcher, S., & MacKinnon, I. (2008). Novelty and diversity in information retrieval evaluation. *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 659–666. doi:10.1145/1390334.1390446

Cober, R. T., Brown, D. A., Levy, P. E., Cober, A. B., & Keeping, L. M. (2003). Organizational Web Sites: Web Site Content and Style as Determinants of Organizational Attraction. *International Journal of Selection and Assessment*, 11(2/3), 158–169. doi:10.1111/1468-2389.00239

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale: Erlbaum.

Coiro, J. (2011). Predicting Reading Comprehension on the Internet. *Journal of Literacy Research*, *43*(4), 352–392. doi:10.1177/1086296X11421979

Coleman, E. B. (1962). Improving comprehensibility by shortening sentences. *Journal of Applied Psychology*, *46*(2), 131.

Costello, A. B., & Osborne, J. W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment, Research & Education*, *10*(7), 1–9. doi:10.1.1.110.9154

Dames, H., Hirschfeld, G., Sackmann, T. & Thielsch, M. T. (under review). Browsing vs. Searching - Exploring the influence of consumers' goal directedness on website evaluation.

De Marsico, M., & Levialdi, S. (2004). Evaluating web sites: exploiting user's expectations. *International Journal of Human-Computer Studies, 60*(3), 381–416. doi:10.1016/j.ijhcs.2003.10.008

De Wulf, K., Schillewaert, N., Muylle, S., & Rangarajan, D. (2006). The role of pleasure in web site success. *Information & Management, 43*(4), 434–446. doi:10.1016/j.im.2005.10.005

DeLone, W. H., & McLean., E. R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of Management Information Systems, 19*(4), 9–30. doi:10.1007/978-1-4419-6108-2

Dinet, J., Chevalier, A., & Tricot, A. (2012). Information search activity: An overview. *Revue Europeene de Psychologie Appliquee, 62*(2), 49–62. doi:10.1016/j.erap.2012.03.004

Douneva, M., Jaron, R. & Thielsch, M.T. (2016). Effects of different website designs on first impressions, aesthetic judgments, and memory performance after short presentation. *Interacting with Computers, 28* (4), 552-567. doi:10.1093/iwc/iwv033

Dumais, S., & Chen, H. (2000). Hierarchical classification of Web content. *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '00*, 256–263. doi:10.1145/345508.345593

Elling, S., Lentz, L., & de Jong, M. (2007). Website Evaluation Questionnaire: Development of a Research-Based Tool for Evaluating Informational Websites. *Electronic Government: 6th International Conference, (EGOV 2007), 4656/2007*, 293–304. doi:10.1007/978-3-540-74444-3

Eysenbach, G., & Köhler, C. (2002). How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *British Medical Journal, 324*(7337), 573–577.

Fillmore, L. (1995). Internet publishing: how we must think. *Journal of Electronic Publishing, 1*(1&2)

Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet information credibility. *Journalism & Mass Communication Quarterly, 77*(3), 515-540.

Flavián, C., Guinalíu, M., & Gurrea, R. (2006). The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information & Management, 43*(1), 1–14. doi:10.1016/j.im.2005.01.002

Fletcher, R. (2006). The impact of culture on web site content, design, and structure: An international and a multicultural perspective. *Journal of Communication Management*, 10(3), 259–273. doi:10.1108/13632540610681158

Fogg, B. J., Kameda, T., Boyd, J., Marshall, J., Sethi, R., Sockol, M., & Trowbridge, T. (2002). *Stanford-makovsky web credibility study 2002: Investigating what makes web sites credible today* (Tech. Rep.). Stanford University: Stanford Persuasive Technology Lab and Makovsky & Company.

Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., ... & Treinen, M. (2001). What makes Web sites credible? A report on a large quantitative study. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 61-68). ACM.

Fogg, B. J., & Tseng, H. (1999). The elements of computer credibility. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 80-87). ACM.

Geißler, H., Donath, T., & Jaron, R. (2003). Von der Schwierigkeit Websites benutzerfreundlich zu gestalten. [The difficulty to design usable websites] *Planung & Analyse*, *30*(2), 42-49.

Hasan, L., & Abuelrub, E. (2011). Assessing the quality of web sites. *Applied Computing and Informatics*, *9*(1), 11-29. doi:10.1016/j.aci.2009.03.001

Hausman, A. V., & Siekpe, J. S. (2009). The effect of web interface features on consumer online purchase intentions. *Journal of Business Research*, *62*(1), 5–13. doi:10.1016/j.jbusres.2008.01.018

Hayes, A. F. (2005, November). *A computational tool for survey shortening applicable to composite attitude, opinion, and personality measurement scales*. Paper presented at the meeting of the Midwestern Association for Public Opinion Research. Chicago, IL.

Hirschfeld, G. & Thielsch, M. T. (2015). Establishing meaningful cut points for online user ratings. *Ergonomics, 58* (2), 310-320. doi:10.1080/00140139.2014.965228

Hirschfeld, G., von Brachel, R. & Thielsch, M. T. (2014). Selecting items for Big Five questionnaires: At what sample size do factor loadings stabilize? *Journal of Research in Personality, 53,* 54-63. doi:10.1016/j.jrp.2014.08.003

Hong, T. (2006). Contributing factors to the use of health-related websites. *Journal of Health Communication, 11*(2), 149–65. doi:10.1080/10810730500526679

Hong, S. Y., & Kim, J. (2004). Architectural criteria for website evaluation - conceptual framework and empirical validation. *Behaviour & Information Technology, 23*(5), 337–357. doi:10.1080/01449290410001712753

Hsieh, T. L., Lo, H. H., Hu, H. H., & Chang, C. C. (2015). The effect of information design on cognitive processing of website navigation. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *25*(5), 548-558. doi: 10.1002/hfm.20568

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. doi:10.1080/10705519909540118

Huizingh, E. (2000). The content and design of web sites: An empirical study. *Information & Management, 37*, 123–134. doi:10.1016/S0378-7206(99)00044-0

ISO (1998) *ISO 9241: Ergonomic Requirements for Office Work with Visual Display Terminals, VDTS—Part 11: Guidance on Usability*. International Organization for Standardisation, Geneva.

ISO (2006). *ISO 9241: Ergonomics of Human-System Interaction – Part 151: Guidance on World Wide Web Interfaces*. Geneva: International Organization for Standardisation.

Johnson, T. J., & Kaye, B. K. (2002). Webelievability: A path model examining how convenience and reliance predict online credibility. *Journalism & Mass Communication Quarterly*, *79*(3), 619–642. doi:10.1177/107769900207900306

Kalyanaraman, S., & Sundar, S. (2006). The psychological appeal of personalized content in web portals: Does customization affect attitudes and behavior? *Journal of Communication*, 56(1), 110–132. doi:10.1111/j.1460-2466.2006.00006.x

Kang, Y., & Kim, Y. (2006). Do visitors' interest level and perceived quantity of web page content matter in shaping the attitude toward a web site? *Decision Support Systems, 42*(2), 1187–1202. doi:10.1016/j.dss.2005.10.004

Karreman, J., van der Geest, T., & Buursink, E. (2007). Accessible website content guidelines for users with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities,* 20, 510–518. doi:10.1111/j.1468-3148.2006.00353.x

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, *46*(5), 604-632. doi:10.1145/324133.324140

Koch, W., & Frees, B. (2016). Dynamische Entwicklung bei mobiler Internetnutzung sowie Audios und Videos [Dynamic development of mobile Internet use plus audio and video]. *Media Perspektiven*, *9/2016*, 418–437.

Kim, S. Y., & Lim, Y. J. (2001). Consumers ' Perceived Importance of and Satisfaction with Internet Shopping. *Electronic Markets*, *11*(3), 148–154. doi:10.1080/101967801681007988

Kim, H., & Niehm, L. S. (2009). The Impact of Website Quality on Information Quality, Value, and Loyalty Intentions in Apparel Retailing. *Journal of Interactive Marketing, 23*(3), 221–233. doi:10.1016/j.intmar.2009.04.009

Kincl, T., & Štrach, P. (2012). Measuring website quality: asymmetric effect of user satisfaction. *Behaviour & Information Technology, 31*(7), 647–657. doi:10.1080/0144929X.2010.526150

Kohli, S., Kaur, S., & Singh, G. (2012). A website content analysis approach based on keyword similarity analysis. *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2012 IEEE/WIC/ACM International Conferences on*, *1*, 254–257. doi:10.1109/WI-IAT.2012.212

Krauss, K. (2003, January). *Testing an e-government website quality questionnaire: a pilot study*. Paper presented at the 5th Annual Conference on World Wide Web Applications (WWW2003). Durban, South Africa.

Kreiner, D. S., Schnakenberg, S. D., Green, A. G., Costello, M. J., & McClin, A. F. (2002). Effects of spelling errors on the perception of writers. *The Journal of general psychology*, *129*(1), 5-17. doi:10.1080/00221300209602029

Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human Computer Studies*, *60*(3), 269–298. doi:10.1016/j.ijhcs.2003.09.002

Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: A methodology for information quality assessment. *Information and Management*, *40*(2), 133–146. doi:10.1016/S0378-7206(02)00043-5

Lehto, T., & Oinas-Kukkonen, H. (2011). Persuasive features in web-based alcohol and smoking interventions: a systematic review of the literature. *Journal of medical Internet research*, *13*(3).

Lin, H.-F. (2007). The Impact of Website Quality Dimensions on Customer Satisfaction in the B2C E-commerce Context. *Total Quality Management & Business Excellence, 18*(4), 363–378. doi:10.1080/14783360701231302

Liu, C., & Arnett, K. P. (2000). Exploring the factors associated with Web site success in the context of electronic commerce. *Information & Management, 38*(1), 23–33.

Liu, C., White, R. W., & Dumais, S. (2010). Understanding web browsing behaviors through Weibull analysis of dwell time. *Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '10* (379-386. doi:10.1145/1835449.1835513

Loiacono, E., Watson, R., & Goodhue, D. (2007). WebQual: An Instrument for Consumer Evaluation of Web Sites. *International Journal of Electronic Commerce, 11*(3), 51–87. doi:10.2753/JEC1086-4415110302

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Journal*, *20*(3), 709–734.

McKinney, V., Yoon, K., & Zahedi, F. (2002). The measurement of Web-customer satisfaction: An expectation and disconfirmation approach. *Information Systems Research, 13*(3), 296–315. doi:10.1287/isre.13.3.296.76

Metzger, M. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology, 58*(13), 2078–2091.

Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics*, *59*, 210–220. doi:10.1016/j.pragma.2013.07.012

Moshagen, M. & Thielsch, M. T. (2010). Facets of visual aesthetics. *International Journal of Human-Computer Studies, 68* (10), 689-709. doi:10.1016/j.ijhcs.2010.05.006

Moshagen, M. & Thielsch, M. T. (2013). A short version of the visual aesthetics of websites inventory. *Behaviour & Information Technology, 32* (12), 1305-1311. doi:10.1080/0144929X.2012.694910

Moustakis, V., Tsironis, L., & Litos, C. (2006). A Model of Web Site Quality Assessment. *The Quality Management Journal, 13*(2), 22–37.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Ozok, A. A., & Salvendy, G. (2001). How consistent is your web design? *Behaviour & Information Technology*, 20(6), 433–447. doi:10.1080/01449290110092260

Palmer, J. W. (2002). Web site usability, design, and performance metrics. *Information Systems Research, 13*(2), 151–167.

Park, Y. A., & Gretzel, U. (2007). Success Factors for Destination Marketing Web Sites: A Qualitative Meta-Analysis. *Journal of Travel Research,* 46(1), 46–63. doi:10.1177/0047287507302381

Parker, M., Moleshe, V., De La Harpe, R., & Wills, G. (2006). An evaluation of Information quality frameworks for the World Wide Web. In *Proceedings of the 8th Annual Conference on WWW Applications*. Bloemfontein, Free State Province, South Africa. Retrieved from http://eprints.ecs.soton.ac.uk/12908/

Rahimnia, F., & Hassanzadeh, J. F. (2013). The impact of website content dimension and e-trust on e-marketing effectiveness: The case of Iranian commercial saffron corporations. *Information and Management, 50*(5), 240–247. doi:10.1016/j.im.2013.04.003

Rains, S. A., & Karmikel, C. D. (2009). Health information-seeking and perceptions of website credibility: Examining Web-use orientation, message characteristics, and structural features of websites. *Computers in Human Behavior, 25*(2), 544–553. doi:10.1016/j.chb.2008.11.005

Ranganathan, C., & Ganapathy, S. (2002). Key dimensions of business-to-consumer web sites. *Information & Management, 39*(6), 457–465. doi:10.1016/S0378-7206(01)00112-4

Robbins, S. S., & Stylianou, A. C. (2003). Global corporate web sites: an empirical investigation of content and design. *Information & Management, 40*(3), 205–212. doi:10.1016/S0378-7206(02)00002-2

Robertson, S., & Zaragoza, H. (2009). The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends in Information Retrieval, 3*(4), 333-389. doi:10.1561/1500000019

Robins, D., & Holmes, J. (2008). Aesthetics and credibility in web site design. *Information Processing & Management*, *44*(1), 386–399. doi:10.1016/j.ipm.2007.02.003

Rosen, D. E., & Purinton, E. (2004). Website design: Viewing the web as a cognitive landscape. *Journal of Business Research*, *57*(7), 787–794. doi:10.1016/S0148-2963(02)00353-3

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods, 1*(2), 199.

Seckler, M., Heinz, S., Forde, S., Tuch, A. N., & Opwis, K. (2015). Trust and distrust on the web: User experiences and website characteristics. *Computers in Human Behavior, 45*, 39–50. doi:10.1016/j.chb.2014.11.064

Selden, S., & Orenstein, J. (2011). Government E-Recruiting Web Sites: The influence of e-recruitment content and usability on recruiting and hiring outcomes in US state governments. I*nternational Journal of Selection and Assessment*, 19(1), 31–40. doi:10.1111/j.1468-2389.2011.00532.x

Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & Herrera-Viedma, E. (2015). Sentiment analysis: A review and comparative analysis of web services. *Information Sciences*, 311, 18–38. doi:10.1016/j.ins.2015.03.040

Shukla, A., Sharma, N. K., & Swami, S. (2010). Website characteristics, user characteristics and purchase intention: mediating role of website satisfaction. *International Journal of Internet Marketing and Advertising, 6*(2), 142. doi:10.1504/IJIMA.2010.032479

Sillence, E., Briggs, P., Harris, P. R., & Fishwick, L. (2007). How do patients evaluate and make use of online health information? S*ocial Science & Medicine, 64*(9), 1853–62. doi:10.1016/j.socscimed.2007.01.012

Smith, A. G. (1997). Testing the surf: criteria for evaluating Internet information resources. *Public Access-Computer Systems Review*, 8(3).

Smith, A. G. (2001). Applying evaluation criteria to New Zealand government websites. *International Journal of Information Management*, *21*(2), 137–149. doi:10.1016/S0268-4012(01)00006-8

Spector, P. E. (1992). *Summated rating scale construction: An introduction* (No. 82). Sage.

Spyridakis, J. H. (2000). Guidelines for authoring comprehensible Web pages and evaluating their success. *Technical Communication, 47*(3), 359–382.

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. *Digital Media, Youth, and Credibility*, 73–100. doi:10.1162/dmal.9780262562324.073

Sutherland, L. A., Wildemuth, B., Campbell, M. K., & Haines, P. S. (2005). Unraveling the web: an evaluation of the content quality, usability, and readability of nutrition web sites. *Journal of Nutrition Education and Behavior,* 37(6), 300–305. doi:10.1016/S1499-4046(06)60160-7

Thielsch, M.T. (2008). *Ästhetik von Websites [Aesthetics of websites]*. Münster: MV Wissenschaft.

Thielsch, M. T., Blotenberg, I. & Jaron, R. (2014). User evaluation of websites: From first impression to recommendation. *Interacting with Computers, 26* (1), 89-102. doi:10.1093/iwc/iwt033

Thielsch, M. T., Engel, R. & Hirschfeld, G. (2015). Expected usability is not a valid indicator of experienced usability. *PeerJ Computer Science*, 1:e19. doi:10.7717/peerj-cs.19

Thielsch, M. T., & Hirschfeld, G. (2012). Spatial frequencies in aesthetic website evaluations–explaining how ultra-rapid evaluations are formed. *Ergonomics*, *55*(7), 731–742. doi:10.1080/00140139.2012.665496

Thielsch, M. T. & Thielsch, C. (2018). Depressive symptoms and web user experience. *PeerJ* 6:e4439. doi:10.7717/peerj.4439

Thielsch, M. T., Träumer, L. & Pytlik, L. (2012). E-Recruiting and fairness – the applicant's point of view. *Information Technology and Management, 13* (2), 59-67. doi:10.1007/s10799-012-0117-x

Thielsch, M. T. & Wirth, M. (2017). Web-based annual reports at first contact: corporate image and aesthetics. *Technical Communication, 64* (4), 282-296.

Thongpapanl, N., & Ashraf, A. R. (2011). Enhance Online Performance Through Website Content and Personalization. *Journal of Computer Information Systems*, *52*(1), 3–13.

Tjur, T. (2009). Coefficients of Determination in Logistic Regression Models—A New Proposal: The Coefficient of Discrimination. *American Statistician*, *63*(4), 366–372. http://doi.org/10.1198/tast.2009.08210

Tsakonas, G., & Papatheodorou, C. (2006). Analysing and evaluating usefulness and usability in electronic information services. *Journal of Information Science, 32*(5), 400–419. doi:10.1177/0165551506065934

Tuch, A. N., Presslaber, E. E., Stöcklin, M., Opwis, K., & Bargas-Avila, J. A. (2012). The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments. *International Journal of Human-Computer Studies*, *70*(11), 794–811. doi:10.1016/j.ijhcs.2012.06.003

Verhagen, T., Boter, J., & Adelaar, T. (2010). The Effect of Product Type on Consumer Preferences for Website Content Elements: An Empirical Study. *Journal of Computer-Mediated Communication*, *16*(1), 139–170. doi:10.1111/j.1083-6101.2010.01536.x

von Brachel, R., Hötzel, K., Hirschfeld, G., Rieger, E., Schmidt, U., Kosfelder, J., … Vocks, S. (2014). Internet-based motivation program for women with eating disorders: eating disorder pathology and depressive mood predict dropout. *Journal of Medical Internet Research, 16*(3), e92. doi:10.2196/jmir.3104

Wang, R. Y., Lee, Y., Pipino, L., Strong, D. (1998). Managing your information as a product. *Sloan Management Review*, *39*(4), 95–106.

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Source Journal of Management Information Systems*, *12*(4), 5–33. doi:10.2307/40398176

Warnick, B. (2004). Online Ethos: Source Credibility in an “Authorless” Environment. *American Behavioral Scientist*, *48*(2), 256–265. doi:10.1177/0002764204267273

Wathen, C. N., & Burkell, J. (2002). Believe it or not: Factors influencing credibility on the Web. *Journal of the American society for information science and technology, 53*(2), 134-144

West, S. G., Finch, J. F., & Curran, P. J. (1995). *Structural Equation Models With Nonnormal Variables: Problems and Remedies* (pp. 56–75). Thousand Oaks: Sage.

Winter, S., & Krämer, N. C. (2012). Selecting Science Information in Web 2.0: How Source Cues, Message Sidedness, and Need for Cognition Influence Users’ Exposure to Blog Posts. *Journal of Computer-Mediated Communication*, *18*(1), 80–96. doi:10.1111/j.1083-6101.2012.01596.x

Yang, Z. L., Cai, S. H., Zhou, Z., & Zhou, N. (2005). Development and validation of an instrument to measure user perceived service quality of information presenting Web portals. *Information & Management, 42*(4), 575–589.

Zhang, X., Keeling, K. B. & Pavur, R. J. (2000). Information quality of commercial web site home pages: An explorative analysis. In *Proceedings of the 21st anniversary international conference on information systems* (pp. 164–175). Atlanta: Association for Information Systems.

Zhao, W., Massey, B., Murphy, J., & Fang, L. (2003). Cultural Dimensions of Website Design and Content. *Prometheus, 21*(1), 74–84. doi:10.1080/0810902032000051027

APPENDIX A. OVERVIEW ON FACETS OF WEBSITE CONTENT IN USER EXPERIENCE RESEARCH

Content facet	Publications covering the specific facet
Accessibility / availability	Abdinnour-Helm et al. (2005), Caro et al. (2008), Karreman et al. (2007), Parker et al. (2006), Ranganathan & Ganapathy (2002), Smith (1997)
Accuracy / adequacy / correctness / consistency / reliability of information	Aladwani & Palvia (2002), Aranyi & van Schaik (2016), Cao et al. (2005), Caro et al. (2008), De Marsico & Levialdi (2004), Hasan & Abuelrub (2011), Moustakis et al. (2006), McKinney et al. (2002), Ozok & Salvendy (2001), Parker et al. (2006), Seckler et al. (2015), Smith (1997), Sutherland et al. (2005), Tsakonas & Papatheodorou (2006), Yang et al. (2005)
Amount of information / data	Caro et al. (2008), Kang & Kim (2006), Palmer (2002), Spyridakis (2000)
broadness / diversity / specificity / variety of information	Agarwal & Venkatesh (2002), Aladwani & Palvia (2002), Caro et al. (2008), Clarke et al. (2008), Hasan & Abuelrub (2011), Palmer (2002), Rosen & Purinton (2004), Selden & Orenstein (2011), Smith (1997)
Completeness / sufficiency	Aladwani & Palvia (2002), Caro et al. (2008), De Wulf et al. (2006), DeLone & McLean (2003), Moustakis et al. (2006), Parker et al. (2006), Smith (1997)
Conciseness of content	Aladwani & Palvia (2002), Caro et al. (2008), Spyridakis (2000)
* Clarity / comprehensibility / ease of understanding / intelligibility / understandability	Aladwani & Palvia (2002), Caro et al. (2008), De Marsico & Levialdi (2004), DeLone & McLean (2003), Parker et al. (2006), Smith (1997), Spyridakis (2000), Thielsch (2008)
* Credibility / authority / believability / reputation / trustworthiness	Caro et al. (2008), De Wulf et al. (2006), Flanagin & Metzger (2000), Fogg & Tseng (1999), Fogg et al., (2001 & 2002), Hasan & Abuelrub (2011), Hong (2006), Loiacono, et al. (2007), Metzger (2007), Parker et al. (2006), Seckler et al. (2015), Smith (1997), Spyridakis (2000), Wathen & Burkell (2002)
Currency / timeliness	Abdinnour-Helm et al. (2005), Agarwal & Venkatesh (2002), Aladwani & Palvia (2002), Caro et al. (2008), De Marsico & Levialdi (2004), De Wulf et al. (2006), Hasan & Abuelrub (2011), McKinney et al. (2002), Parker et al. (2006), Seckler et al. (2015), Smith (1997), Sutherland et al. (2005), Tsakonas & Papatheodorou (2006)
* Informativeness	Chakraborty et al. (2005), Hausman & Siepke (2009), Kang & Kim (2006), Lin (2007), Rahimnia & Hassanzadeh (2013), Shukla et al. (2010)
Interactivity / responsiveness / support	Caro et al. (2008), Kalyanaraman & Sundar (2006), Park & Gretzel (2007)
* Likability / attractiveness / entertainment	Caro et al. (2008), Huizingh (2000), Kang & Kim (2006), Thielsch (2008)
Novelty	Caro et al. (2008), Clarke et al. (2008), Kalyanaraman & Sundar (2006)
Objectivity	Caro et al. (2008), Hasan & Abuelrub (2011), Parker et al. (2006)
Originality / uniqueness of content	Aladwani & Palvia (2002), Moustakis et al. (2006), Smith (1997)
* Perceptions of specific content, e.g., information on procedures, organizational culture, feedback, etc.	Braddy et al. (2009), Caro et al. (2008), Cober et al. (2003), Selden & Orenstein (2011), Thielsch, Träumer & Pytlik (2012)
Personalization / tailored information	DeLone & McLean (2003), Kalyanaraman & Sundar (2006), Moustakis et al. (2006), Loiacono et al. (2007), Park & Gretzel (2007), Thongpapanl & Ashraf (2011)
* Relevance	Agarwal & Venkatesh (2002), Cao et al. (2005), Caro et al. (2008), De Wulf et al. (2006), DeLone & McLean (2003), Hasan & Abuelrub (2011), Hong (2006), Kalyanaraman & Sundar (2006), McKinney et al. (2002), Parker et al. (2006), Spyridakis (2000), Tsakonas & Papatheodorou (2006)
Security / perceived security / privacy	Caro et al. (2008), Casaló et al. (2007), DeLone & McLean (2003), Lin (2007); Park & Gretzel (2007), Parker et al. (2006), Ranganathan & Ganapathy (2002)
* Usefulness / utility of content / value added	Aladwani (2002), Aladwani & Palvia (2002), Aranyi & van Schaik (2016), Caro et al. (2008), Hong & Kim (2004), Loiacono et al. (2007), McKinney et al. (2002), Moustakis et al. (2006), Parker et al. (2006), Tsakonas & Papatheodorou (2006), Yang et al. (2005)

Note: Overview of different website content facets as researched in prior publications. Facets that are best suited for subjective survey based evaluations are marked with an *.

APPENDIX B: ITEMS AND ITEM STATISTICS

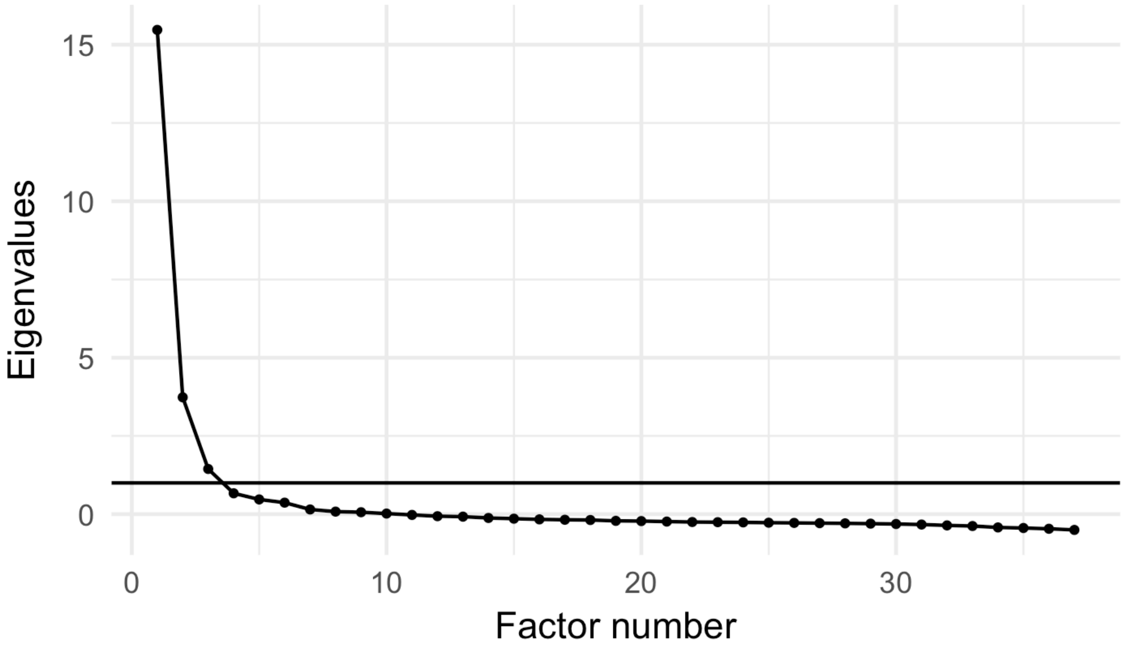
B.1 Original items analyzed in study 1.

Itemnumber	Facet	Item in German	Item in English	Item source
01	Clarity	Die Informationen sind exakt.	The Information is precise.	Krauss, 2003; Thielsch, 2008; Yang et al., 2005
02*	Clarity	Die Inhalte sind anschaulich aufbereitet.	The contents of the website are clearly presented.	Geißler et al., 2003; Thielsch, 2008
03	Clarity	Die einzelnen Sätze sind einfach zu lesen.	The individual sentences are easy to read.	De Wulf et al., 2006; Geißler et al., 2003; Thielsch, 2008
04*	Clarity	Der Sprachgebrauch in den Texten ist geläufig und allgemein verständlich.	The language used in the texts is current and easy to understand.	(Elling et al., 2007); Geißler et al., 2003; (Smith, 2001); Thielsch, 2008
05	Clarity	Die Inhalte auf der Website sind gut erklärt.	The contents on the website are well explained.	(Elling et al., 2007); Thielsch, 2008
06	Clarity	Die Informationen auf der Website sind in sich schlüssig.	The information on the website is coherent.	(Rosen & Purinton, 2004)
07	Clarity	Die Informationen auf der Website sind fehlerfrei.	The information on the website is accurate.	Ahn et al., 2007; Aladwani, 2002; Cao et al., 2005; Chakraborty et al., 2005; De Wulf et al., 2006; Hong & Kim, 2004; (Lin, 2007)
08	Credibility	Ich werde auf der Website objektiv informiert.	I get informed objectively on this website.	Geißler et al., 2003; (Hong & Kim, 2004); Thielsch, 2008
09	Credibility	Die Informationen sind zuverlässig.	The information is reliable.	Ahn et al., 2007; Aladwani & Palvia, 2002; Thielsch, 2008
10	Credibility	Die Informationen auf der Website sind überzeugend.	The information on the website is convincing.	(Choi & Rifon, 2002)
11*	Credibility	Ich finde die auf der Website dargebotenen Informationen glaubwürdig.	I find the information provided on the website to be authentic.	Chakraborty et al., 2005; De Wulf et al., 2006; (Zhang et al., 2000)
12*	Credibility	Ich kann den Informationen auf der Website vertrauen.	I can trust the information on the website.	(Cao et al., 2005); (De Wulf et al., 2006)
13	Credibility	Die auf der Website dargebotenen Informationen sind sachlich.	Information on the web site is objective.	Hong & Kim, 2004
14*	Credibility	Die auf der Website dargebotenen Informationen sind seriös.	The information provided on the website is reliable.	(Cao et al., 2005); (De Wulf et al., 2006)
15	Credibility	Die auf der Website dargebotenen Informationen sind unparteiisch.	The information on the website is unbiased.	Hong, 2006; (Smith, 2001)
16	Informativeness	Ich kann mich über alles informieren, das mich interessiert.	I can get information on anything I am interested in.	Geißler et al., 2003; Thielsch, 2008
17*	Informativeness	Die Informationen sind qualitativ hochwertig.	The information is of high quality.	Cao et al., 2005; Kim & Lim, 2001; Thielsch, 2008
18	Informativeness	Die Website liefert mir die benötigten Informationen.	The website provides me with the required information.	Abdinnour-Helm et al., 2005; Thielsch, 2008,
19*	Informativeness	Die Website ist informativ.	The website is informative.	Kang & Kim, 2006; Karreman et al., 2007; Shukla et al., 2010; Thielsch, 2008
20	Informativeness	Die Website beinhaltet reichhaltige Informationen.	The website contains extensive information.	(Palmer, 2002); Lavi & Tractinsky, 2004
21*	Likeability	Die Website weckt mein Interesse.	The website arouses my interest.	Agarwal & Venkatesh, 2002; Thielsch, 2008

22	Likeability	Der Inhalt der Website gefällt mir.	I like the content of the website.	Thielsch, 2008
23	Likeability	Die Website ist unterhaltsam.	The website is enjoyable.	Kang & Kim, 2006; Thielsch, 2008
24*	Likeability	Ich lese diese Website gerne.	I enjoy reading the website.	Thielsch, 2008
25*	Likeability	Die Inhalte der Website sind spannend.	The contents of the website are exciting.	Thielsch, 2008
26	Likeability	Das Lesen der Website macht Spaß.	Reading the website is fun.	(Cao et al., 2005); (De Wulf et al., 2006); (Kang & Kim, 2006)
27	Relevance	Ich erhalte die Informationen, die ich erwarte.	I get the information I expect.	Thielsch, 2008
28	Relevance	Die Website beinhaltet alle relevanten Informationen.	The website contains all relevant information.	Aladwani, 2002; Chakraborty et al., 2005; Thielsch, 2008; Zhang et al., 2000
29	Relevance	Die Inhalte der Website sind wichtig.	The contents of the website are important.	Thielsch, 2008
30	Relevance	Themen, die auf der Website angesprochen werden, bedeuten mir persönlich viel.	Issues addressed on the website mean a lot to me.	Thielsch, 2008
31	Relevance	Die Texte auf der Website laden zum Lesen ein.	The texts on the website stimulate further reading.	Geißler et al., 2003; Thielsch, 2008
32	Originality / Uniqueness of content	Die Inhalte der Website sind anregend.	The contents of the website are inspiring.	Thielsch, 2008
33	Originality / Uniqueness of content	Der Inhalt der Website weckt mein Interesse.	The content of the website sparks my interest	(De Wulf et al., 2006)
34	Originality / Uniqueness of content	Die Websiteinhalte motivieren mich, die Seite wieder zu besuchen.	The content of the website motivates me to revisit the site.	Thielsch, 2008
35	Originality / Uniqueness of content	Die Inhalte der Website sind so wichtig, dass ich sie mir ausdrucken oder speichern würde.	Contents of the website seem so important to me, that I would print or save them.	Geißler et al., 2003; Thielsch, 2008
36*	Usefulness	Ich finde die Informationen auf der Website sind nützlich.	I find the information on the website to be useful.	Aladwani, 2002; Cao et al., 2005; (Elling et al., 2007); (Lin, 2007); Thielsch, 2008
37*	Usefulness	Die Texte liefern mir kurz und bündig die wichtigsten Informationen.	The texts provide me information in a clear and concise manner.	Geißler, Donath & Jaron, 2003; Thielsch, 2008
38	Usefulness	Von der Website kann man etwas lernen.	One can learn from this website.	Kang & Kim, 2006
39	Usefulness	Die Inhalte der Website sind professionell.	The contents of the website are professional	(Smith, 2001)
40	Usefulness	Die auf der Website dargebotenen Informationen sind ausreichend.	The information provided on the website is sufficient.	Abdinnour-Helm et al., 2005; De Wulf et al., 2006; Elling et al., 2007

Note: If item source is given in parentheses the item was not directly taken from this source but adapted based on it. An asterisk at the item number indicates selected items for the final Web-CLIC questionnaire.

B.2 Scree plot resulting from exploratory factor analysis of study 1.



Note: The straight line illustrates an eigenvalue of 1; N = 1226.

B.3 Items, factor loadings, means, and standard deviations for remaining 37 items in exploratory factor analysis of study 1.

An asterisk at the item number indicates selected items for the final Web-CLIC questionnaire.

Item number	Likeability	Credibility	Clarity	Informativeness	*M*	*SD*
32	**0.695**	-0.005	0.152	0.079	3.097	1.530
21*	**0.847**	0.004	-0.017	0.085	2.894	1.613
33	**0.862**	0.040	-0.081	0.083	2.943	1.677
22	**0.713**	0.012	-0.032	0.271	3.212	1.617
24*	**0.786**	0.034	0.114	-0.007	2.617	1.453
30	**0.705**	-0.095	-0.258	0.226	2.501	1.638
31	**0.581**	-0.029	**0.337**	-0.027	3.148	1.613
34	**0.730**	0.041	-0.014	0.122	2.640	1.636
25*	**0.820**	0.019	-0.021	0.023	2.865	1.524
26	**0.795**	0.035	0.297	-0.226	2.812	1.520
23	**0.629**	-0.073	**0.328**	**-0.338**	2.883	1.545
08	0.023	**0.544**	0.216	-0.001	3.763	1.665
11*	0.037	**0.962**	-0.005	-0.104	4.772	1.425
12*	0.030	**0.983**	-0.014	-0.132	4.500	1.425
13	-0.004	**0.889**	-0.158	0.069	4.292	1.678
14*	0.008	**0.861**	-0.034	0.024	4.761	1.446
15	-0.013	**0.657**	0.017	-0.164	3.805	1.611
27	-0.089	-0.048	**0.523**	**0.358**	4.538	1.565
02*	0.219	0.061	**0.645**	-0.197	3.900	1.737
03	-0.022	-0.144	**0.737**	-0.073	5.113	1.555
37*	0.012	0.078	**0.594**	0.076	4.103	1.626
04*	-0.014	-0.074	**0.598**	0.010	5.294	1.449
05	0.063	0.027	**0.619**	0.144	4.250	1.483
18	-0.058	0.007	**0.330**	**0.595**	4.079	1.653
36*	0.297	0.115	-0.105	**0.610**	3.936	1.723
29	**0.415**	-0.021	-0.155	**0.563**	3.575	1.749
20	-0.033	0.077	0.091	**0.638**	4.509	1.670
19*	0.107	0.231	-0.004	**0.525**	4.607	1.537
17*	0.135	**0.306**	0.049	**0.440**	3.877	1.522
01	-0.101	0.278	**0.314**	**0.354**	4.028	1.410
28	-0.157	0.070	**0.396**	**0.490**	4.061	1.573
09	-0.038	**0.389**	0.290	0.204	4.002	1.402
06	-0.095	0.075	**0.495**	**0.334**	4.591	1.394
10	0.119	0.273	0.244	**0.370**	4.136	1.542
39	0.059	0.280	**0.345**	0.156	4.321	1.645
40	-0.176	**0.384**	0.271	0.210	4.497	1.573
38	**0.309**	0.106	-0.036	**0.378**	3.911	1.756

Note: *M* = mean, *SD* = standard deviation; loadings higher than .3 are marked bold; N = 1226.

APPENDIX C: STIMULI

C.1 Experts ratings for websites in pre-study of study 1

Ratings of content quality are ordered from good to bad, websites marked bold were selected for study 1. Websites with familiarity values above 10 %, or content quality and overall impression ratings influenced by age or gender (as indicated by significant correlations) were excluded from the final set of study 1.

Website URL	Website category	Familiarity	Content quality		Overall grade	
			M	*SD*	*M*	*SD*
http://www.travian.de	Entertainment	13.64%	5.11	(0.83)	2.22	(0.81)
http://www.sprengsatz.de	**Information site**	**8.00%**	**5.00**	**(0.95)**	**2.38**	**(0.74)**
http://www.tognum.com	Corporate website	4.35%	4.95	(1.20)	2.64	(1.09)
http://www.vag-armaturen.de	**Corporate website**	**0.00%**	**4.95**	**(1.39)**	**2.47**	**(1.31)**
http://www.hotel-blog.de	Information site	4.35%	4.80	(1.15)	2.65	(0.81)
http://www.mvjob.de	**E-recruiting**	**4.35%**	**4.76**	**(1.48)**	**2.67**	**(1.02)**
http://www.marsh.de	**Corporate website**	**0.00%**	**4.59**	**(1.18)**	**2.95**	**(1.09)**
http://www.scienceticker.info	Information site	12.50%	4.42	(1.64)	2.68	(1.20)
http://www.pricerunner.de	E-commerce	4.17%	4.41	(1.37)	3.05	(1.09)
http://www.deutsche-allgemeine-zeitung.de	**Information site**	**0.00%**	**4.35**	**(1.31)**	**3.00**	**(1.26)**
http://www.girlsgogames.de	**Entertainment**	**4.17%**	**4.29**	**(1.23)**	**3.05**	**(1.02)**
http://www.lynet.de	**Corporate website**	**0.00%**	**4.26**	**(1.05)**	**3.26**	**(0.99)**
http://www.preistester.de	**E-commerce**	**9.52%**	**3.88**	**(1.17)**	**3.41**	**(1.00)**
http://www.assistenz.org	**E-recruiting**	**0.00%**	**3.86**	**(1.31)**	**3.67**	**(1.24)**
http://www.playzo.de	Entertainment	8.70%	3.84	(0.96)	3.37	(0.90)
http://www.szene.it	**Web portal**	**0.00%**	**3.65**	**(1.15)**	**3.87**	**(0.92)**
http://www.finanztreff.de	Information site	16.00%	3.58	(1.22)	3.63	(0.90)
http://www.neopreis.de	E-commerce	4.17%	3.50	(1.14)	3,77	(0.87)
http://www.excite.de	Web portal	27.27%	3.00	(1.31)	4.13	(1.06)

Note: Level of familiarity was assessed dichotomously (known/unknown), content quality on a seven-point Likert scale (ranging from 1 = “very bad” to 7 = “very good”), and overall grade on a six-point grading scale (ranging from 1 = “very good” to 6 = “insufficient”). N = 37; due to dropout, each website was rated by n = 15 to n = 25 experts. Screenshots can be requested via the corresponding author.

C.2 URLs of websites tested in study 2, 3, and 4.

Website category	Definition of category	Website URLs study 2	Website URLs study 3	Website URLs study 4
Download & Software	Websites providing free or fee-based apps, programs or codes for downloads.	http://www.freeware-download.com/ http://www.softwareload.de/	http://www.freeware.de	
E-Commerce	Websites with the primary aim of buying and selling.	http://www.buch.de http://www.danto.de/ http://www.karstadt.de/	http://www.stylepit.de	http://www.preistester.de
E-Learning	Online learning content and webpages for learning.	http://www.fahrschuleonline.de/ http://www.fit-fuer-den-aufschwung.de/	http://www.sgd.de	
E-Recruiting & E-Assessment	Web-based recruiting and assessment.	http://www.absolventa.de/ http://jobboerse.arbeitsagentur.de/	http://www.jobware.de	http://www.assistenz.org www.mvjob.de
Entertainment	Websites with the main aim to entertain	http://www.clipfish.de/ http://www.onlinegames.de/		http://www.girlsgogames.de
Information site	Websites with a strong focus on information (also containing passive use of weblogs and wikis).	http://dict.leo.org/ http://www.ftd.de/ http://www.tagesschau.de/ http://www.taz.de/ http://www.zeit.de	http://www.handelsblatt.com MedOnline (mock site)	http://www.deutsche-allgemeine-zeitung.de http://www.sprengsatz.de
Presentation & Self-portrayal (corporate websites)	Websites of institutions, organizations, and companies for representation and image cultivation	http://www.bmw.de/de/de/index.html http://www.brueninghoff.de/ http://www.dp-dhl.com/de http://www.meuter.de/ http://www.originalhaflingerpferde-deutschland.de	http://www.kpmg.com	http://www.lynet.de http://www.marsh.de http://www.vag-armaturen.de
Search engines	Websites serving for the search of other websites, products, services or the like.	http://de.ask.com http://www.bing.com/	http://www.ixquick.com	
7Web portals	Websites providing an overview of many different issues, offering information and additional links and services.	http://www.deutschland.de/ http://www.einfach-teilhaben.de		http://www.szene.it
Weblogs and Social Sharing	Websites serving for creation of virtual chronological diaries, collaborative text editing, immediate networking and interaction of the users or for sharing of resources (e.g. pictures, links, video)	http://www.basicthinking.de http://www.blog.de/ http://www.flickr.com http://www.kopfschuettel.de/ http://www.mister-wong.de/		

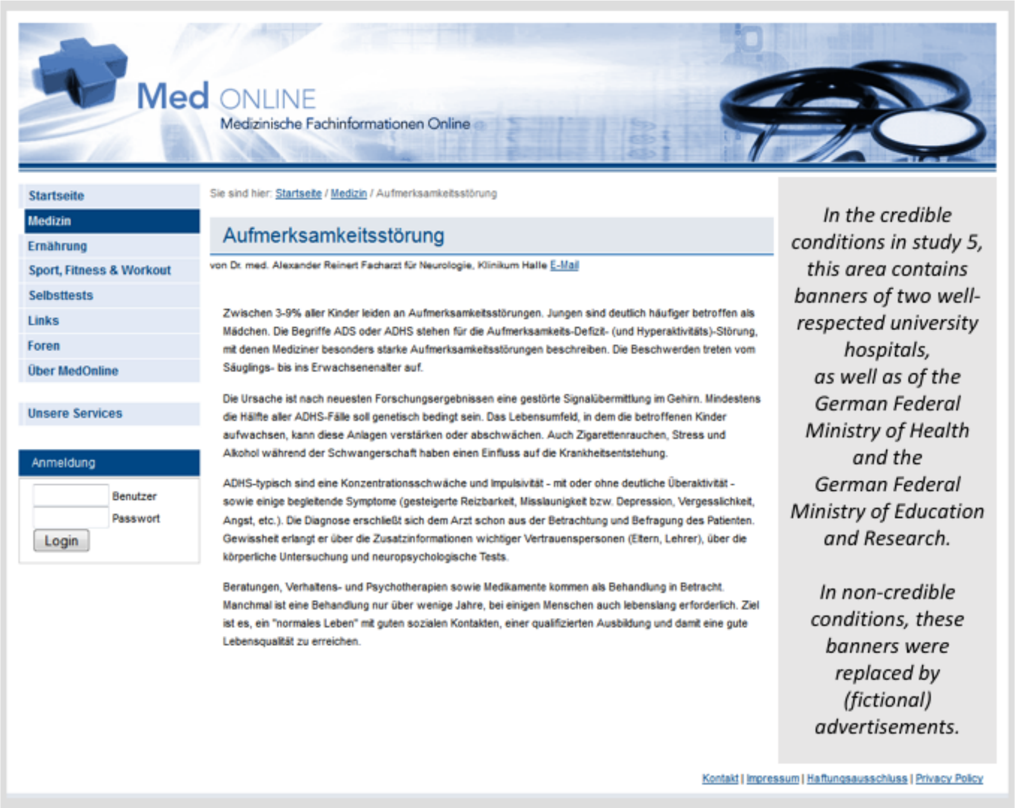
Note. In study 2, 3, and 4 fully-functional websites were linked with the named URL, screenshots can be requested via the corresponding author.

C.3 Examples for treatments and stimuli used in study 5

Manipulated facet	Text examples
High clarity	Between 3 and 9 % of all children suffer from attention deficit disorder. Boys are significantly more often affected than girls. The terms ADD or ADHD stand for the attention deficit (and hyperactivity) disorder, with which physicians describe especially heavy attention deficit disorders. Grievances occur from infancy to adulthood. According to latest research results, the cause is a defected signal transmission in the brain. At least half of all ADHD cases are supposed to be genetically determined. The living environment, which the affected children grow up in, can aggravate or attenuate these dispositions. Smoking cigarettes, stress and alcohol during pregnancy influence the development of the disease. (…)
Low clarity	The prevalence of attention deficit disorders, which boys are more affected by than girls, is 3 to 9% among children, whereupon the terms ADD and ADHD represent the attention deficit (and hyperactivity) disorder, which physicians use to describe especially heavy attention deficit disorders, whose grievances occur from infancy to adulthood. According to latest research results, the cause is a defected signal transduction in the brain, whereupon at least half of all ADHD cases are supposed to be genetically determined and the living environment, which the affected children grow up in, can aggravate or attenuate these dispositions, which already entails the influence of tobacco consumption, stress and alcohol during pregnancy on the pathogenesis. (…)
High informativeness	Between 3 and 9 % of all children suffer from attention deficit disorder. Boys are significantly more often affected than girls. The terms ADD or ADHD stand for the attention deficit (and hyperactivity) disorder, with which physicians describe especially heavy attention deficit disorders. Grievances occur from infancy to adulthood. According to latest research results, the cause is a defected signal transmission in the brain. At least half of all ADHD cases are supposed to be genetically determined. The living environment, which the affected children grow up in, can aggravate or attenuate these dispositions. Smoking cigarettes, stress and alcohol during pregnancy influence the development of the disease. (…)
Low informativeness	Some children suffer from attention deficit disorders. In some cases, a drug that has chemical similarities to speed is used for treatment. Incidentally, speed is not the same as crystal. However, crystal is a substance that has similarities to speed. Such a substance similar to speed was first produced in 1887 by the chemist L. Edeleanu at the Humboldt University of Berlin. That is where Edeleanu wrote his doctoral thesis from 1883 to 1887 under the supervision of August Wilhelm von Hofmann. Hofmann married four times during his lifetime. However, three of his wives died young. He had eleven kids. He died in 1892 and was buried at the cemetery of Dorotheenstadt. Later, the sale of speed was restricted in many countries. (…)
High credibility	Attention deficit disorder *by Dr. med. Alexander Rainert, neurology specialist, Clinic of Halle, E-Mail* Between 3 and 9 % of all children suffer from attention deficit disorder. Boys are significantly more often affected than girls. The terms ADD or ADHD stand for the attention deficit (and hyperactivity) disorder, with which physicians describe especially heavy attention deficit disorders. Grievances occur from infancy to adulthood. (…)
Low credibility	Attention deficit disorder *by DJAlex71* Bewteen 3 and 9 % of al children suffer from attention defficit disorder. Boys are significantly more often afected than girls. The terms ADD or ADHD stand for the attention defficit (and hyperactiviti) disorder, with whitch physicians describe especially Heavy attention defficit disorders. Grievances occurer from infancy to Adulthood. (…)

Note. Original text manipulations were performed in German, displayed texts in this table are illustrations of manipulations.

Screenshot example from study 5



Sie sind hier: Startseite / Medizin / Aufmerksamkeitsstörung

Aufmerksamkeitsstörung

von Dr. med. Alexander Reinert Facharzt für Neurologie, Klinikum Halle E-Mail

Zwischen 3-9% aller Kinder leiden an Aufmerksamkeitsstörungen. Jungen sind deutlich häufiger betroffen als Mädchen. Die Begriffe ADS oder ADHS stehen für die Aufmerksamkeits-Defizit- (und Hyperaktivitäts)-Störung, mit denen Mediziner besonders starke Aufmerksamkeitsstörungen beschreiben. Die Beschwerden treten vom Säuglings- bis ins Erwachsenenalter auf.

Die Ursache ist nach neuesten Forschungsergebnissen eine gestörte Signalübermittlung im Gehirn. Mindestens die Hälfte aller ADHS-Fälle soll genetisch bedingt sein. Das Lebensumfeld, in dem die betroffenen Kinder aufwachsen, kann diese Anlagen verstärken oder abschwächen. Auch Zigarettenrauchen, Stress und Alkohol während der Schwangerschaft haben einen Einfluss auf die Krankheitsentstehung.

ADHS-typisch sind eine Konzentrationsschwäche und Impulsivität - mit oder ohne deutliche Überaktivität - sowie einige begleitende Symptome (gesteigerte Reizbarkeit, Misslaunigkeit bzw. Depression, Vergesslichkeit, Angst, etc.). Die Diagnose erschließt sich dem Arzt schon aus der Betrachtung und Befragung des Patienten. Gewissheit erlangt er über die Zusatzinformationen wichtiger Vertrauenspersonen (Eltern, Lehrer), über die körperliche Untersuchung und neuropsychologische Tests.

Beratungen, Verhaltens- und Psychotherapien sowie Medikamente kommen als Behandlung in Betracht. Manchmal ist eine Behandlung nur über wenige Jahre, bei einigen Menschen auch lebenslang erforderlich. Ziel ist es, ein "normales Leben" mit guten sozialen Kontakten, einer qualifizierten Ausbildung und damit eine gute Lebensqualität zu erreichen.

In the credible conditions in study 5, this area contains banners of two well-respected university hospitals, as well as of the German Federal Ministry of Health and the German Federal Ministry of Education and Research.

In non-credible conditions, these banners were replaced by (fictional) advertisements.

Note. Banners in the right area of the screenshot had to be removed due to copyright restrictions.

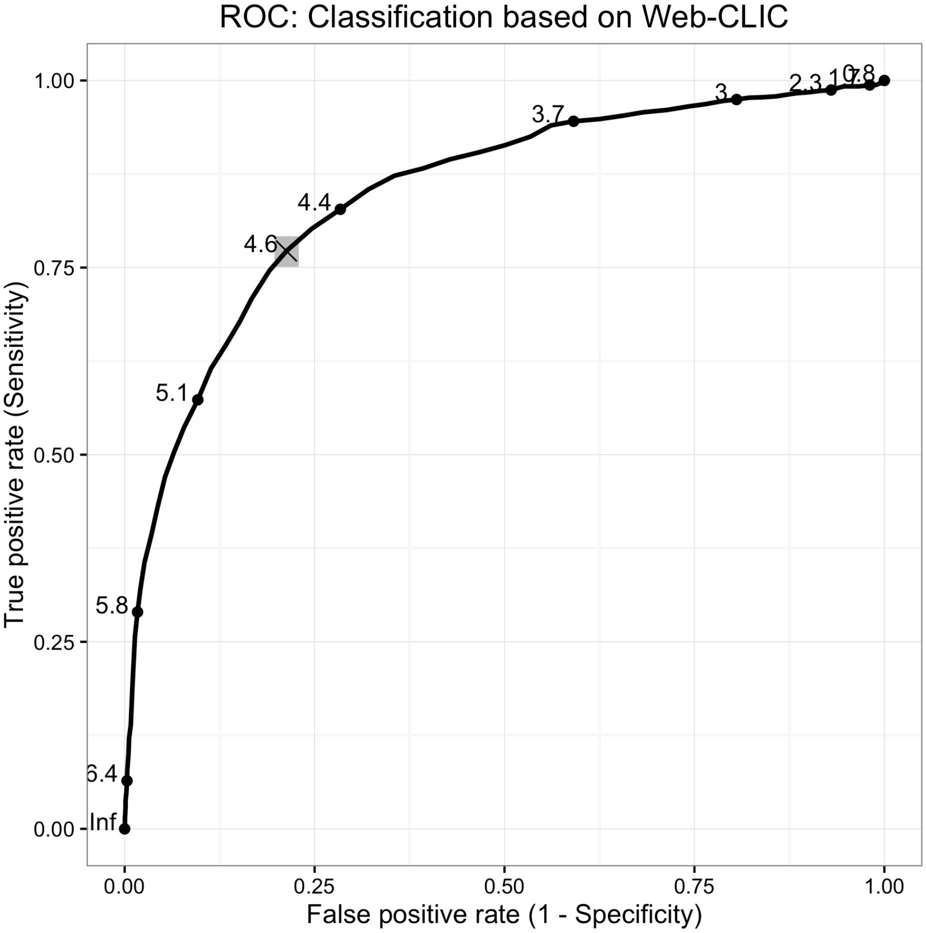
APPENDIX D: BENCHMARK AND CUT POINT ANALYSES

D.1 Correlations between age, educational level and the Web-CLIC scores

	WEB-CLIC sum score	Clarity	Likeability	Informativeness	Credibility
Age (N ≥ 5336)	-.001	-.021	.047**	-.011	-.025
Education level (N ≥ 4275)	-.067**	-.062**	-.075**	-.063**	-.026

Note: Differences in sample size are caused by missing demographic data; ** = *p* < .01

D.2 ROC curve for the Web-CLIC against the dichotomous good versus unattractive rating



Supplement:

The Web-CLIC questionnaire in German

Bitte beurteilen Sie den Inhalt der Ihnen vorliegenden Website anhand der folgenden Aussagen auf einer Skala von 1 (stimme gar nicht zu) bis 7 (stimme voll zu). Vielen Dank!

	Stimme gar nicht zu	*Stimme nicht zu*	*Stimme eher nicht zu*	*neutral*	*Stimme eher zu*	*Stimme zu*	*Stimme voll zu*
Die Inhalte sind anschaulich aufbereitet.	①	②	③	④	⑤	⑥	⑦
Die Texte liefern mir kurz und bündig die wichtigsten Informationen.	①	②	③	④	⑤	⑥	⑦
Der Sprachgebrauch in den Texten ist geläufig und allgemein verständlich.	①	②	③	④	⑤	⑥	⑦
Die Website weckt mein Interesse.	①	②	③	④	⑤	⑥	⑦
Die Inhalte der Website sind spannend.	①	②	③	④	⑤	⑥	⑦
Ich lese diese Website gerne.	①	②	③	④	⑤	⑥	⑦
Die Informationen sind qualitativ hochwertig.	①	②	③	④	⑤	⑥	⑦
Ich finde die Informationen auf der Website sind nützlich.	①	②	③	④	⑤	⑥	⑦
Die Website ist informativ.	①	②	③	④	⑤	⑥	⑦
Die auf der Website dargebotenen Informationen sind glaubwürdig.	①	②	③	④	⑤	⑥	⑦
Die auf der Website dargebotenen Informationen sind seriös.	①	②	③	④	⑤	⑥	⑦
Ich kann den Informationen auf der Website vertrauen.	①	②	③	④	⑤	⑥	⑦

Supplement:

The Web-CLIC questionnaire in English

Please judge the content of present website according to the following statements on a scale ranging from 1 (strongly disagree) to 7 (strongly agree). Thank you very much!

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
The contents of the website are clearly presented.	①	②	③	④	⑤	⑥	⑦
The texts provide me information in a clear and concise manner.	①	②	③	④	⑤	⑥	⑦
The language used in the texts is current and easy to understand.	①	②	③	④	⑤	⑥	⑦
The website arouses my interest.	①	②	③	④	⑤	⑥	⑦
The contents of the website are exciting.	①	②	③	④	⑤	⑥	⑦
I enjoy reading the website.	①	②	③	④	⑤	⑥	⑦
The information is of high quality.	①	②	③	④	⑤	⑥	⑦
I find the information on the website to be useful.	①	②	③	④	⑤	⑥	⑦
The website is informative.	①	②	③	④	⑤	⑥	⑦
I find the information provided on the website to be authentic.	①	②	③	④	⑤	⑥	⑦
The information provided on the website is reliable.	①	②	③	④	⑤	⑥	⑦
I can trust the information on the website.	①	②	③	④	⑤	⑥	⑦