

Linek, Stephanie B.

Article

Order Effects in Usability Questionnaires

Journal of Usability Studies

Suggested Citation: Linek, Stephanie B. (2017) : Order Effects in Usability Questionnaires, Journal of Usability Studies, ISSN 1931-3357, Usability Professionals' Association, Bloomingdale, Ill., Vol. 12, Iss. 4, pp. 164-182

This Version is available at:

<http://hdl.handle.net/11108/312>

Kontakt/Contact

ZBW – Leibniz-Informationszentrum Wirtschaft/Leibniz Information Centre for Economics
Düsternbrooker Weg 120
24105 Kiel (Germany)
E-Mail: info@zbw.eu
<https://www.zbw.eu/de/ueber-uns/profil-der-zbw/veroeffentlichungen-zbw>

Standard-Nutzungsbedingungen:

Dieses Dokument darf zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden. Sie dürfen dieses Dokument nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Sofern für das Dokument eine Open-Content-Lizenz verwendet wurde, so gelten abweichend von diesen Nutzungsbedingungen die in der Lizenz gewährten Nutzungsrechte.

Terms of use:

This document may be saved and copied for your personal and scholarly purposes. You are not to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. If the document is made available under a Creative Commons Licence you may exercise further usage rights as specified in the licence.

Order Effects in Usability Questionnaires

Stephanie B. Linek

Researcher, Post-Doc
ZBW–Leibniz Information
Centre for Economics
Duesternbrooker Weg 120
24105 Kiel
Germany
s.linek@zbw.eu

Abstract

Usually, websites consist of several components (e.g., a homepage and subsections) that can be very different from each other. Thus, it can be advantageous to assess the usability separately for each part. Hence the question arises if and how the order in which the usability evaluation was done influences the results.

The presented empirical study investigated order effects for the arrangement of the usability evaluation of the components of a website. The use case was the website of a library that included the homepage and three online services. For two of the services, the association with the website owner was obvious. For one service, the association was weak; that is, the connection with the website owner was not obvious. The independent variable was the order of the usability evaluation; that is, the homepage was either evaluated before or after the services. The measurement instrument was the System Usability Scale that was applied for a retrospective evaluation.

There was a significant order effect for the weakly associated service: The usability was rated better if the rating was done after the evaluation of the homepage. The order effect can be best explained as a halo effect that emerges because a good image of the website owner was made cognitively available by the preceding evaluation of the homepage. Theoretical and practical implications are discussed.

Keywords

usability questionnaire, order effects, halo effect, part-whole effect, usability evaluation of websites, homepage, online services



Introduction

For today's established companies or public institutions (e.g., universities, libraries), a modern and continuously updated website is a must. Usually, a website does not consist of a single page but rather of a homepage with general information as well as specific subpages or special online services. These components are often designed to address different issues and (at least partly) different target groups. In this context, it is important to bear in mind that the components of a website can be very dissimilar from each other. Additionally, website updates often only target one component (e.g., a specific online service) and not the whole website. In such constellations, it is advantageous to conduct differentiated usability evaluations: an evaluation of the homepage itself and a separate evaluation for each of the incorporated services and other components.

At first sight, this seems trivial, and the obvious solution is to apply multiple scales in usability evaluations, that is, a separate usability scale for the homepage and for each of the essential components. However, on second glance, the question arises of how this should be managed, and especially, in which order the usability scales for the homepage and the other components should be arranged. To answer this question, I conducted an empirical study to investigate possible order effects for the arrangement of the usability evaluation of the components of a website. The use case was the usability evaluation of the website of a Library 2.0, namely the ZBW (www.zbw.eu). The ZBW is the world's largest research infrastructure for economic literature, either offline or online. The website of the ZBW not only consists of a homepage with regular information about the ZBW, it also offers several online services. (A detailed description follows in the section on the use case of the study.)

The following theoretical section explores the theoretical background and provides an overview on order effects in questionnaires with special focus on the so-called halo effect and the part-whole effect. Subsequently, the use case and the research questions of the study are described. The methodological section presents the design of the study, the participants, the description of the questionnaire, and the practical procedure of data assessment. Afterwards the results of the study are reported. The discussion provides an interpretation of the results, their theoretical and practical implications, and an outlook for future research. This paper closes with tips for usability practitioners that can be derived from the presented findings.

Theoretical Background

This section presents the literature for using questionnaires to evaluate usability, the effect of completing a questionnaire on a participant's cognitive processing, and a discussion of the halo effect and the part-whole effect.

Evaluating Usability by Questionnaires

In addition to qualitative usability studies that aim to identify concrete usability problems, quantitative indicators are also needed. This is because they have several advantages like objectivity and the possibility of scientific generalization (Nunnally, 1975; Sauro & Lewis, 2016). From a practical point of view, quantitative indicators enable the monitoring of a website's usability and make improvements measurable (Linek & Tochtermann, 2011). Besides objective quantitative indicators like success rate or time on task, subjective quantitative indicators in the form of questionnaires with rating scales can also be helpful. Questionnaires with standardized scales are a method commonly used to assess subjective quantitative usability indicators. Nowadays, several standardized usability scales are available: both short ones and detailed longer evaluations scales. For a longer more detailed usability evaluation, the ISONORM (Prümper, 1999) or the Isometrics (Gediga, Hamborg, & Düntsch, 1999) scales are popular. A very popular short scale for usability evaluation is the System Usability Scale (SUS; Brooke, 1996). Bangor, Kortum, and Miller (2009) pointed out the subjective meaning of SUS scores and also discussed the aspects of validity and reliability of the SUS. Tullis and Stetson (2004) and Berkman and Karahoca (2016) described a comparison with other questionnaires for assessing usability of websites. Other short usability questionnaires include, for example, the Computer System Usability Questionnaire (CSUQ: <http://hcibib.org/perلمان/question.cgi>) by Lewis (1995), the Questionnaire for User Interaction Satisfaction (QUIS: <http://www.lap.umd.edu/QUIS/index.html>) by Chin, Diehl, and Norman (1988), and the

Usability Metric for User Experience (UMUX) by Finstad (2010; see also Berkman and Karahoca, 2016).

The study described in this paper used the SUS as measurement instrument, and thus, I want to provide some details about the scale. The SUS consists of 10 items in the form of statements that have to be rated on a 5-point Likert-scale ranging from "strongly agree" to "strongly disagree." According to Bentler and Chou (1987), variables that were measured by a Likert-scale with five or more gradings can be conceptualized as a metric scale, and thus, from a purely statistical perspective, the values of the SUS could be treated like interval data. However, this does not necessarily mean that the interpretation of mean SUS scores (of studies with broader samples) follows the same logic. Recent work on the interpretation of SUS scores (Lah & Lewis, 2016; Sauro & Lewis, 2016) indicated that mean SUS scores should be practically interpreted as a curved grading scale. Based on the data of 241 usability studies, Sauro and Lewis (2016) developed a curved grading scale that used the percentile range of SUS scores as a basis for the practical interpretation in form of grades A to F (see Table 8.5 in Sauro and Lewis on p. 204). The median (50th percentile) is set as the average grade (C) because it is middle of the range. The highest and lowest 15 percentile points represent grades A and F, respectively. The curved grading scale describes a more suitable interpretation of the SUS values because it is based on percentile ranges and grades. This also allows a practical interpretation of mean differences. According to the described curved grading scale whether the mean difference has practical relevance or not depends on the absolute value. For example, a difference of 15 points at the very top of the scale (e.g., 100 versus 85) has no practical relevance because both can be categorized as grade A+. However, the same difference of 15 points can have a very high practical relevance for other absolute values, for example, 75 versus 60 because these values correspond to grade B (75) versus grade D (60). Values below 51.6 can be interpreted as an F grade and even very large differences (e.g., between 10 and 40) have no practical relevance because the usability is unacceptable anyway.

Lah and Lewis (2016) used this curved grading scale to investigate the effect of expertise on SUS scores. Persons with higher experience (in relation to computer use and the tested software) not only showed higher success rates, lower error rate, and shorter time on task but also reported higher SUS scores for the tested software. The findings of Lah and Lewis (2016) and the results of similar research (Borsci, Federici, Bacci, Gnaldi, & Bartolucci, 2015; Kortum & Johnson, 2013; Lewis, Utesch, & Maher, 2015; McLellan, Muddimer, & Peres, 2012) imply that the prior knowledge of the users should be considered when interpreting overall SUS scores. A software product that received excellent ratings from a group of experts might have only an average or even low perceived usability for a group of novices.

Questionnaires and the Logic of Conversation

From a psychological point of view, the act of completing a questionnaire can be seen as a form of cognitive information processing. The questions have to be understood and interpreted before the answers can be given. Accordingly, the "questions shape the answer" (Schwarz, 1999). The underlying principle is the logic of conversation, namely Grice's cooperative principle (Grice, 1975). The logic of every successful conversation is that the sender (speaker) gives us meaningful information (i.e., new, understandable information, etc.) and the receiver (listener) tries to make sense of the given information. In this sense-making process, the context plays a decisive role. For a questionnaire, the "context" of the conversation is not only the instruction, but also the layout, the presentation (e.g., online or paper-based), and the arrangement (e.g., the order) of the questions.

A very prominent example of a context effect is the social desirability effect (Ahammer, 1971; Edwards, 1957), that is, the tendency to behave in a socially desirable way and meet the expectations of the reference group or the conversational partner. In questionnaires, this effect is well-known as socially desirable responding (overview by Paulhus, 2002). Other context effects may arise due to the order of questions. So-called order effects are mainly known and investigated in relation to learning, memory, and the identification of information. Prominent examples include the primacy and the recency effect (Atkinson & Shiffrin, 1968; Murdock, 1962).

The literature on the methodology of questionnaires also identified various order effects (Schuman & Presser, 1981). Order effects in questionnaires relate to the phenomenon whereby

a prior question can influence the answers to subsequent questions. Thus, the order of questions is usually permuted and statistically controlled. In relation to the evaluation of a website with its components, especially the halo effect and the part-whole effect are of relevance. In the following sections I will describe these two order effects in more detail.

Halo Effect

Order effects in questionnaires can be due to the priming of information (Bargh & Pietromonaco, 1982; Strack, Martin, & Schwarz, 1988). If specific information is primed by a preceding question, this information is more cognitively available for the person when answering the subsequent questions and thus, might influence the subsequent answers. For example, if a person is first asked about the drawbacks of nuclear power, this informational priming could influence the subsequent judgement of a non-profit organization like Greenpeace. One prominent manifestation of priming is the so-called halo effect, whereby one characteristic also influences the perception of other characteristics (Thorndike, 1920). A popular example refers to the attractiveness stereotype: Beautiful people are perceived as nice and intelligent (Dion, Berscheid, & Walster, 1972). In this context, even a single word can create a halo effect. For instance, using "fruit sugar" instead of "sugar" on the ingredient list increases the perceived healthfulness of food (Sütterlin & Siegrist, 2015). Similarly, priming can also lead to a halo effect in questionnaires: If a preceding question makes a desirable (or undesirable) attribute of the evaluated object visible, this can influence the subsequent answers on other attributes. For example, if a person is first asked about a company's social engagement, this can influence his or her subsequent rating of the trustworthiness of the company.

Generally, the halo effect is also a well-known phenomenon in usability evaluation, due to the relation between the perceived aesthetics (i.e., beauty of the surface design) and the perceived usability. In the seminal article "What Is Beautiful Is Usable," Tractinsky, Katz, and Ikar (2000) described the relation between design aesthetics and perceived usability. Even though most studies found a significant association between design aesthetics and usability judgements, there were also contradictory findings. Besides the "what is beautiful is usable" perspective, some studies also support the "what is usable is beautiful" notion (Tuch, Roth, Hornbæk, Opwis, & Bargas-Avila, 2012). Additionally, there have also been findings that an aesthetic design made users more aware of usability problems (Murphy, Stanney, & Hancock, 2003).

The inconsistent findings imply that the relation between aesthetics and perceived usability is complex and other sources of influence can moderate the relationship. For example, Hartmann, Sutcliffe, and De Angeli (2008) proposed a framework on the relationship between quality judgements (including usability and aesthetics) and user background, tasks, and content. The empirical studies on their framework not only demonstrated that quality criteria influence each other (in the sense of a halo effect) but also revealed the influence of the user's needs and motivation. A similar argument can be found by Aljukhadar and Senecal (2015) who referred to the technology acceptance model (TAM) proposed by Davis (1989) to explain the perceived ease of use as a result of the complex interplay between the quality of a website, its interactivity, and its aesthetics. Additionally, there were empirical results that the user's experience also influences the perceived usability (Berkman & Karahoca, 2016; Lah & Lewis, 2016; McLellan, Muddimer, & Peres, 2012) and goodness, whereas the perceived aesthetics remains stable over time (Hassenzahl, 2004). This is very similar to the findings of Raita and Oulasvirta (2010) that the users' expectations before the actual use moderate the subsequent usability judgements.

In relation to a user's expectations, one important source of influence might be the image of the owner of the website (the organization or person). Empirical research provided several examples for how image exerts an influence: The institutional prestige of a university can increase the licensing rate (Sine, Shane, & Di Gregorio, 2003), the reputation of an organization can be protective during crisis management (Coombs & Holladay, 2006), and the brand equity of real products can cause a halo effect (Leuthesser, Kohli, & Harich, 1995).

Based on these findings, the question arises of whether and how the image of the website owner (visible by the logo or the headline of the website) influences the perception of the usability of the website. This is even more interesting in relation to the different components of a website, because the website owner could be more or less visible for each component. The reasoning is as follows: Normally, the website owner (person or organization) and its image are

highly visible on the homepage because the usual purpose of the homepage is to present the website owner and the related image. Thus, most of the content describes the organization (or person) or is related to the organization (or person) that is behind the website. The degree of association with the website owner can differ for the various online services of a website. The association is relatively clear if the website owner is obvious for the users of a service (hereinafter referred to as *strongly associated service*). A strong association can be established due to different characteristics. For example, the *surface design* can create a clear association if the service is directly integrated in the homepage and shares the same layout (e.g., integrated live chat or online help). The *purpose of the service* can also lead to a clear association if the functionality of the service is directly connected with the purpose of the website (e.g., the online reservation service of a hotel website or the literature search service of a library website). However, for some services the connection can be less obvious (hereinafter referred to as *weakly associated services*): This can be the case if the service is not directly embedded in the website but takes the form of a separate external environment with a different layout. Similarly, for an unusual extra-service (e.g., the babysitter agency service of a diving school or the publishing service of a library), the association with the website owner is also (probably) rather weak.

For strongly associated services it should not matter if the researcher first asks the respondent to evaluate the homepage or the services because in each case the association with the website owner is evident (and thus, the website owner and the related image are highly cognitively salient). However, for weakly associated services, the question arises of whether the preceding evaluation of the homepage can influence the evaluation of weakly associated services. That means, when one first asks respondents to evaluate the homepage, the website owner and the related image could be primed and thereby could influence the subsequent evaluation of the service in the form of a halo effect (of the primed image). Thus, for weakly associated services the order of the evaluation of the homepage and the service should influence the usability evaluation.

Part-Whole Effect

Another important order effect in questionnaires is the so-called part-whole effect (Mason, Carlson, & Tourangeau, 1994; Schwarz, Strack, & Mai, 1991): If several questions belong to the same conversational context, one can differentiate between general questions (whole) and specific questions (parts). The order of the general versus the specific questions can influence the results in the form of a part-whole effect. An often-cited example is a survey on life happiness: The general question might be "How happy are you with life in general?" and the specific questions could be "How happy are you with your job (with dating, with your family, etc.)?". If one first asks the question on happiness with the job and afterwards the question on general happiness, then the general answer reflects general happiness with life besides the job. In contrast, if one first asks the question on general happiness, the answer will also include happiness with the job.

A part-whole effect can arise in the form of an assimilation effect or in the form of a contrast effect. Depending on the constellation of the general and the specific questions as well as the contextual framing of the questions, Schwarz, Strack, and Mai (1991) made the following predictions for part-whole effects:

- If there is only one specific question, the part-whole effect should be a *contrast effect*, because of Grice's cooperative principle and the maxim of relation that states that the conversational partner should give relevant (new) information. That means if the specific question was presented first, the answer to the subsequent general question would include the remaining information (in the example above, without reference to happiness with the job). This in turn leads to a lower correlation (or no correlation) compared to the correlation of the reverse order of the two questions (i.e., general question in the first place).
- If there are multiple specific questions, the part-whole effect should be an *assimilation effect*. That means if the general question is presented after the multiple special questions, the general answer is a kind of summary, which means that the correlation is higher compared to the inverse order (i.e., asking the general question first).

It is important to note that the part-whole effect can possibly influence the means, but does not necessarily lead to mean differences. Rather a part-whole effect influences the correlation between the general and the specific questions. This means the critical indicator for a part-whole effect is a significant difference between the correlations (of the two different orders).

When conducting usability evaluations of websites, it is an open question how the user perceives the homepage and the other components. If the homepage is seen as the “whole” (because it is the general starting point and represents the website owner), the other subsections and integrated services might be perceived as the “parts.” Consequently, the ordering of the evaluation could lead to a part-whole effect.

Use Case and Research Questions

The following sections present the description of the use case and the research questions used in this study.

Description of the Use Case

For this study, I investigated the order effects in usability questionnaires. The use case was the website of a Library 2.0, namely the ZBW. As described in the introduction, the ZBW is the world’s largest research infrastructure for economic literature. The website of the ZBW not only consists of the homepage (<http://www.zbw.eu/en/>) but also offers several online services. The three main online services are the following:

- The online help “EconDesk” (<http://www.zbw.eu/en/service/reference-desk/>) is where customers can request information about literature and economic topics by email, phone, or chat.
- The literature search service “EconBiz” (<http://www.econbiz.de/>) has international literature and information. The literature search service is the most popular service of the ZBW and represents the very core of a library in the sense of providing literature.
- The publishing portal “EconStor” (<http://www.econstor.eu/dspace/?locale=en>) is where customers can search within the publications of other customers and can (under certain conditions) publish their own manuscripts. The publishing portal is a rather special service and less popular among the regular customers because it mainly addresses economists with more scientific experience (who have also written papers by themselves that they want to publish).

The investigation of order effects was part of a larger and more detailed usability evaluation of the ZBW’s website conducted in preparation for a complete redesign. The detailed evaluation served as a benchmark for comparison (with evaluations after the redesign) and also provided heuristic insights for the planned redesign. In this contribution, we only report the results on order effects. The redesign of the ZBW website has since been completed. The linkages in the text above lead to the new website design. Figures 1, 2, 3, and 4 show screenshots of the original design that was the basis of this empirical study.

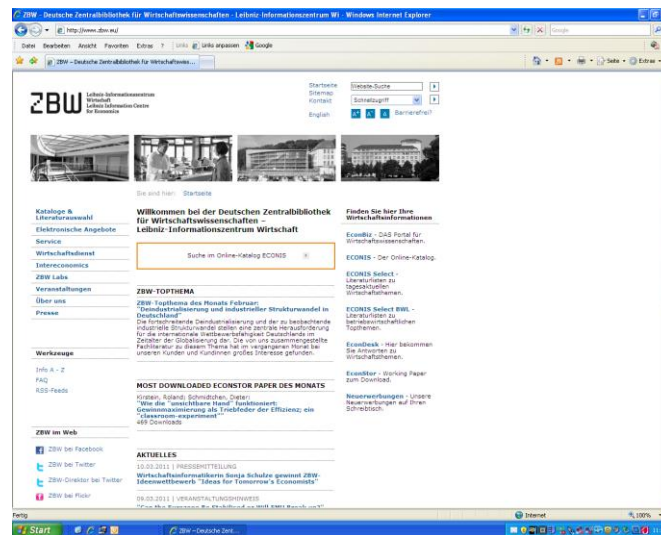


Figure 1. Screenshot of the homepage of the former website of the ZBW.



Figure 2. Screenshot of the online help "EconDesk" of the former website of the ZBW.

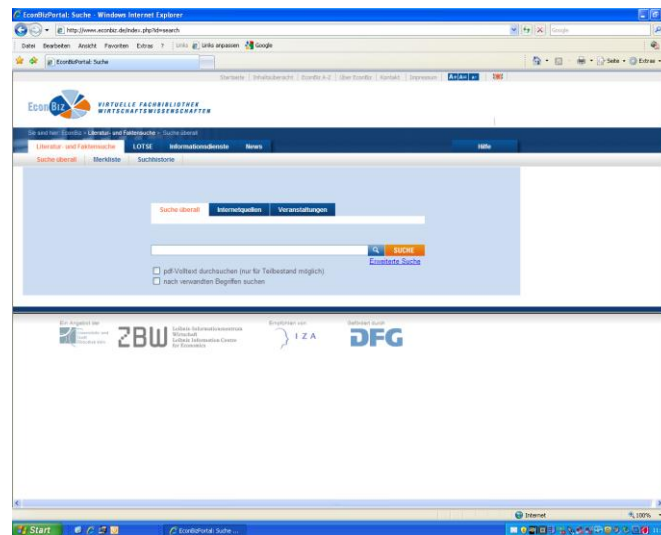


Figure 3. Screenshot of the literature search service “EconBiz” of the former website of the ZBW.



Figure 4. Screenshot of the publishing portal “EconStor” of the former website of the ZBW.

Even though the three services were the ZBW’s main services, some were more visibly associated with the ZBW’s website; this was due to the linkage with the website, the layout, and the core purpose of the services. To estimate the strength of the association of the services with the website, I used the following heuristics:

- A more immediate linkage strengthens the association because this makes the connection between the service and the website more obvious: The association is strengthened if the service is directly integrated in the website, and the association is weakened if the service is implemented as a separate environment.
- A shared layout of the website and the service strengthens the association because it makes it visually apparent that there is a common owner.

- A prominent announcement or a high popularity of a service strengthens the association because the importance of the service for the website is more obvious.
- Having a main functionality that reflects the character of the website strengthens the association. If the service reflects the core purpose of the website, the connection is logically strengthened by the use itself.

The application of these heuristics to the three services can be described as follows: For the *online help*, the connection was obvious because this service was directly integrated into the website and shared the same layout. The *literature search service* was not directly embedded in the homepage but implemented as an external environment. However, it was announced and linked on the homepage (starting page) as the core functionality of the Library 2.0. Additionally, the literature search service was the ZBW's most popular service and reflects the nature of the ZBW as a Library 2.0 (searching and finding literature). The third service, the *publishing portal*, was implemented as an external environment with a different layout. It was announced and linked on the homepage, but in contrast to the literature search it was a rather unusual service for a library and was less well-known by the regular online visitors of the ZBW. Thus, here the association between the ZBW's website and the publishing portal was less pronounced than with the online help and the literature search service. Based on these described differences I made the following assumptions about the varying degree of the association of the three services with the website:

- The online help qualifies as a strongly associated service due to the visible surface design (because it was directly integrated in the website and shared the same layout).
- The literature search service qualifies as a strongly associated service due to its core purpose and popularity (because it was the most popular service that was announced on the homepage and reflected the nature of the organization).
- The publishing portal qualifies as a weakly associated service (because it had a different surface design, was implemented as an external environment, and had a core purpose that was rather unusual for a website related to a library).

Research Questions

I measured the usability of the homepage and the three services by separate evaluation scales for each of them. The measurement instrument was the System Usability Scale (SUS; Brooke, 1996). The order of the usability evaluation (by means of separate SUS form sheets) for the homepage and for each of the services was systematically permuted. The permutations provided the data for the analysis of order effects. Hence, the investigation of order effects was twofold: First, from a practical point of view, I compared the SUS means. This comparison tested if and how the order of the usability evaluation influenced the absolute ratings of usability. This comparison also enables analysis of a possible halo effect. Second, from a more theoretical point of view, I specifically tested for a possible part-whole effect by means of additional correlative analyses. The two research questions of the study were as follows:

- How does the order of the evaluation (homepage and services) affect the means of the evaluation results?
- Are there specific order effects, namely a halo effect or a part-whole effect?

The use case was a realistic one and the attributes of the services were not systematically manipulated; in fact, I characterized the given services from a theoretical and descriptive point of view. Thus, I did not consider formulating testable hypotheses. However, due to the considerations listed in sections on the halo effect and part-whole effect and on the description of the use case, I derived the following predictions for possible order effects:

The halo effect: For the concrete use case of a Library 2.0, the website owner (ZBW) could be described as having an image as a highly trustworthy and innovative institution. In this sense, the ZBW could be seen as a "precious label" that might cause a halo effect for the subsequent evaluations of the services, that is, there should be higher usability ratings for the services if the homepage was evaluated before them (i.e., condition HP-pre) compared to the inverse order (evaluating the homepage after the services, i.e., condition HP-post). This effect should be most pronounced for the weakly associated service (publishing portal). For the two strongly associated services, there was an additional open question, namely whether the initial evaluation of the homepage could lead to a halo effect or if the image of the website owner was

already present due to the inherently strong association. Additionally, it was an open question if the association due to the surface design (online help service) or the association due to the purpose (literature search service) would lead to different effects.

Part-whole effect: There were two possible predictions for a part-whole effect.

- Multiple specific questions (three services) could lead to an assimilation effect for all services: At first glance, the separated evaluation of the homepage and the three services reflected a constellation of a questionnaire with multiple special questions (three services) and one general question (homepage) in the same context, that is, website of a library. Thus, a possible part-whole effect should be there in the form of an assimilation effect for all three services, that is, the correlation between the general and the special questions should be higher in the case that the general question was asked after the special question (i.e., correlation higher in the condition HP-post).
- Only the essential literature search service (which reflects the nature of a Library 2.0) could be seen as a "part" of the "whole" (homepage) and thus, there should be a contrast effect for the literature search but not for the other services: It could also be the case that not all of the services are seen as "parts" of the "whole" (i.e., homepage). In this relation, a strong association could be seen as a necessary prerequisite but not a sufficient condition for a part-whole effect. Rather, the service had to be seen as an essential part of the homepage. This was the case for the literature search service because it was typically for a library. However, a very general service like the online help or a rather unusual service like the publishing portal might not be seen as "parts." Thus, in the sense of a part-whole effect there was only one special question (literature search service) and one general question (homepage). Therefore, there should be a contrast effect for the literature search service only (and no significant effects for the online help and the publishing service).

Methodology

The following sections present the design of the study, the participants, and the description of the questionnaire and the procedure used in this study.

Study Design

The *independent variable* was the order of the evaluation of the homepage and the services. This resulted in a between-design with two groups (two independent groups each with different participants): Evaluation of the homepage before the services (HP-pre) versus evaluation of the homepage after the services (HP-post).

The *dependent variables* were the means of the SUS scores for the homepage (HP-SUS) and the three services, that is, the online help (OH-SUS), the literature search service (LS-SUS), and the publishing service (PP-SUS).

Please note that the usability evaluation by the SUS was based on the prior experiences of the participants. It was a pure survey study and no usability tasks were presented immediately before the evaluation. The application of the SUS without direct tasks is not the traditional manner, but as Grier, Bangor, Kortum, and Peres (2013) pointed out, a retrospective evaluation can be beneficial if a product or service is complex and feature rich or has a customizable interface. In such cases, it is very difficult to present a representative task set that is in line with the individual use of the customers. A retrospective evaluation instead comprises the integrated individual user experiences with the product or service. In the particular use case of the ZBW website, most components can be qualified as rather complex: The homepage offered many different possibilities (reflected by a very extensive navigation menu). Furthermore, the literature search service was feature rich and comprised a broad variety of options like specific filters and connections to other data bases. Similar, the publishing service could be used in quite different ways (i.e., not only for publishing but also for searching) and was designed for more experienced scientists that were familiar with the specialist jargon. Only the online help service was rather simple. Thus, considering the complexity of the homepage, the literature search service, and the publishing service, a retrospective evaluation without preceding tasks seems justifiable.

For the *control variables*, I assessed the following variables: sociodemographics (age, gender, family status), education and occupation, profession or field of study and semester, experience with scientific work, prior experience with the ZBW and its services, experience with computers and the Internet, experience with professional search via the Internet (frequency, duration, and self-rated competence). For reasons of statistical control, the order of the evaluation of the services was also systematically permuted for each of the two experimental conditions. Neither the assessed control variables nor the additional permutations of the services influenced the results on order effects.

Participants

When selecting participants, I aimed to recruit the core target group of the ZBW's customers because of the practical background of the study. The prerequisites for participation were that the participants already knew and used the ZBW's homepage as well as the literature search service. Because the other two services were less popular and less essential, participants did not have to be familiar with the online help and publishing portals. These prerequisites for participation corresponded with the majority of the usual ZBW clients. The participants were recruited personally in the buildings (reading room) of the ZBW in Hamburg and Kiel. Additionally, flyers and mouth-to-mouth recruitment were used.

The sample contained 47 male and 49 female participants that were equally distributed across the two experimental conditions (HP-pre: 23 males and 25 females; HP-post: 24 males and 24 females). The age of the participants was at average 27.33 years and varied between 19 and 53 years. Most of the participants were students in different fields of economics (80%), a smaller portion was employed (12%), or employed and a student (8%). Altogether 42% of the participants were also doing scientific work.

All participants ($n = 96$) were familiar with the homepage and the literature search service because this was the prerequisite for participation. However, only a smaller sample of the participants knew the online help service ($n = 37$; 39%) and the publishing service ($n = 30$; 31%), so the subsamples for the analysis of the OH-SUS and the PP-SUS were smaller.

Description of the Questionnaire

The following sections present the SUS and the questionnaire types used in this study.

Measurement instrument for the dependent variables

The measurement instrument for the dependent variables was the System Usability Scale (SUS). As described in the theoretical section, the SUS is a short scale with 10 items in the form of statements. The statements were modified with respect to the object of evaluation (e.g., instead of "software" I inserted "EconBiz" or "the homepage of the ZBW"). Each statement had to be rated on a Likert scale from 1 to 5. The SUS sum-score was calculated as described in Brooke (1996). The possible sum-score was between 0 and 100. A sum-score of 0 indicated very low usability; a sum-score of 100 reflected excellent usability.

Composition of the complete questionnaire

As already mentioned, the complete questionnaire was not only comprised of the SUS but also included several additional questions that were needed for further insight (for internal librarians' purposes). Most of the additional questions were presented after the SUS. None of the additional questions influenced the results. (The interested reader can contact the author for a full version with the concrete wording in German.)

A first draft of the questionnaire was presented to a small sample of six people. All six people were employees of the ZBW and familiar with the online services. Their feedback was solely used to test the comprehensibility and appropriateness (in relation to the services) of the wording. Based on these pilot tests, some slight changes in the wording of questions were made. However, the standardized scales for usability evaluation (SUS) were not changed (even when criticized by the test persons).

The questionnaire started with a short introduction and the assessment of sociodemographic variables. Afterwards the assessment of the usability of the homepage and the services using the SUS took place. Immediately before each SUS form sheet was presented, a screenshot of the evaluation object (homepage or one of the services) was presented as a visual reminder.

The screenshots were identical with the screenshots shown in Figures 1, 2, 3, and 4. The order of the SUS form sheets for the homepage and the three services was systematically permuted. (Even though some participants knew only one service, they were shown each of the SUS form sheets of the three services, but could indicate if they had never used the service in question. That means all participants received multiple specific questions, but eventually answered not all of them.) Additionally, the literature search service was also evaluated using the ISONORM (because this was the core service of the library and the librarians wanted a more detailed usability evaluation). After the usability evaluation of the homepage and the three services was completed, the additional questions and tasks that had been compiled for internal librarian purposes were presented. This included the assessment of the content quality of the literature search service by the SQuaLL (Linek, 2015), a scribbling task (like described in Linek & Tochtermann, 2015), and several control variables and open questions for additional information. (Please note that I presented no specific usability tasks. The scribbling task included only a playful marking and commenting of printed-out screenshots with different colors. Additionally, the scribbling task took place after the usability assessment was completed.)

Procedure

The participants worked in group sessions in a separate room in the ZBW buildings. The duration of the complete test sessions varied between 30 minutes and 90 minutes. As an incentive, each participant received a €20 voucher for a popular online shop. After a short welcome, the instructor informed the participants that their data would be treated anonymously and strictly confidentially. In order to avoid politeness effects, the instructor explicitly stated that the participants should not give polite answers. Instead, the instructor explained that honest and open answers were needed as a basis for improvements. The participants were instructed (orally and in written form) to work on the questions in the given order. During the whole study, the instructor was present. The questionnaire was handed out and answered in a written paper-pencil based format. During the test sessions, some sweets were offered. After the completion of the questionnaire, the instructor took the filled-out materials and conducted a short post-interview with each participant. Subsequently, they received the voucher as a reward for participation and had again the opportunity to ask questions or make comments. The instructors requested that they not talk about the study for the next few weeks.

Results

The analyses of order effects comprised two steps: First, I tested the SUS values for mean differences in dependency of the order of the questions (comparison of the groups HP-pre versus HP-post); this allowed for the analysis of a possible halo effect and other order effects that might influence the general level of the usability evaluation. Second, I calculated correlative analyses for the investigation of a possible part-whole effect.

Testing Order Effects for the Means of the Usability Evaluation

Overall, the usability evaluation of the homepage and the three services was rather high. In the process, the homepage received the lowest values and the online help received the best evaluation (from the smaller sample that was familiar with it). According to the curved grading scale by Sauro and Lewis (2016, p. 204) the overall score (all) of the homepage (74.40) and the publishing portal (75.83) corresponded with grade B; the overall score of the online help (78.18) and the literature search (77.21) corresponded with grade B+. The descriptive values of the usability evaluation are listed in Table 1.

Table 1. Means (m), Standard Deviations (s), and Number of Valid Cases (n) for the Two Independent Groups: Homepage First (HP-pre) Versus Services First (HP-post)

Evaluation scale	Condition	m	s	n
HP-SUS (homepage)	HP-pre	73.85	14.90	48
	HP-post	74.95	19.96	48
	All	74.40	17.53	96
OH-SUS (online help)	HP-pre	79.64	12.59	14
	HP-post	77.28	18.86	23
	All	78.18	16.61	37
LS-SUS (literature search)	HP-pre	75.16	16.25	48
	HP-post	79.27	11.05	48
	All	77.21	13.98	96
PP-SUS (publishing portal)	HP-pre	83.00	8.77	15
	HP-post	68.67	12.50	15
	All	75.83	12.87	30

The influence of the independent variable, that is, the order of the evaluation of the homepage and the services (general versus specific), on the means of the SUS was analyzed using a one-way ANOVA (comparison of the two independent groups HP-pre versus HP-post).

The analyses revealed a significant order effect for the evaluation of the publishing service ($F = 13.221$; $p = .001$): If the publishing portal was evaluated after the homepage, then the usability of the publishing portal was rated as significantly higher than the evaluation before the homepage. The found difference between 83.00 versus 68.67 corresponded with different grades according to the curved grading scale by Sauro and Lewis (2016, p. 204). While the mean of the condition HP-pre (83.00) corresponded with grade A, the mean of the condition HP-post (68.67) indicated grade C. Thus, the found significant difference also provided evidence for a meaningful difference in the subjective user experience.

Additional control analysis: As described in the section on methodology not all participants were familiar with the online help and the publishing portal because these two services were less popular. Because the publishing portal was only known by about one-third of the sample, it might be the case that the order effect found here is due to a different answering behavior of the smaller selective portion of people who knew and used the publishing portal. As mentioned in the description of the use case, the publishing portal is a very specific service that is mainly used by more scientifically experienced users who had already written scientific texts. Thus, the subsample probably had much more prior knowledge with scientific information search and research-related services. If the higher prior experience (or some other selective characteristics) of this subsample is the (hidden) reason for the order effect found here, then a separate analysis of this subsample might lead to analogous order effects for the other two services. Thus, I tested if this subsample also showed a different answering behavior for the other services that might be obscured by the answers of the other two-thirds of the participants of the complete sample. For this purpose, I analyzed the subsample separately. However, for this subsample as well, there was only a significant order effect for the publishing portal ($F = 13.221$; $p = .001$) but not for the other services or the homepage. This analogous pattern of results provided the first evidence that the found order effect is not dependent on the participants' prior experience.

Analysis of the Part-Whole Effect: Correlations Between the Usability Indices

As pointed out in the theoretical section, a part-whole effect does not necessarily influence the means of the answers. Rather, the correlations between the general and the specific questions are the critical criteria. Thus, in order to investigate possible part-whole effects, I calculated the correlations of the general question (evaluation of the homepage, i.e., HP-SUS) with each of the special questions (evaluation of the services, i.e., OH-SUS, LS-SUS, and PP-SUS) for the two independent groups HP-pre and HP-post.

The correlations for the two independent groups (HP-pre versus HP-post) were statistically analyzed by Fischer Z-transformation and the Z-test (two-tailed test because of the two alternative predictions on a possible assimilation-effect versus a contrast-effect). The correlations, the values of Fischer's Z and the values of the Z-test, are listed in Table 2.

Table 2. Comparisons of Correlations: Correlations (r) Between the Evaluation of the Homepage and the Evaluation of the Single Services, Standardized Values (Fischer Z) of the Correlations, and Number (n) of Valid Cases.

Service	HP-pre	HP-post	Fischer's Z-test
Online help	r = .427 Fischer Z = .456 n = 14	r = .703 Fischer Z = .873 n = 23	Z = -1.111 p = .267
Literature search	r = .473 Fischer Z = .514 n = 48	r = .164 Fischer Z = .165 n = 48	Z = 1.653 p = .099
Publishing portal	r = .198 Fischer Z = .201 n = 15	r = .510 Fischer Z = .563 n = 15	Z = -0.887 p = .374

Note. The values of the Z-test (Z and p) for comparison are listed in the right-hand column.

There were no significant differences between the pairs of correlations. Thus, the analyses provided no statistical support for a part-whole effect.

Discussion and Conclusion

I found a significant order effect for the means of the usability evaluation: The usability of the weakly associated service was estimated better if it was rated after the evaluation of the homepage. The correlative analyses revealed that this was not a part-whole effect (i.e., no significant difference between the correlations of the publishing portal and the homepage for the groups HP-pre and HP-post). Taken together, the results matched the predictions made for a halo effect: Due to the weak association of the publishing portal with the website owner, the image of the organization (ZBW) was by itself not cognitively salient during the usability evaluation. However, when the participants were first asked to evaluate the homepage, the organization "ZBW" was primed. This in turn made salient that the publishing service was a service of the ZBW. That means that if the homepage was evaluated before the publishing portal, the image of the ZBW was made cognitively available and the previously weak association between the ZBW and the publishing portal changed and became a stronger association. Thus, the order effect for the (weakly associated) publishing portal could best be interpreted as a halo effect of the "precious" label ZBW. That means the good image of the ZBW resulted in a halo effect for the subsequent evaluation of the publishing portal. For the other two services, the connection of the ZBW was cognitively available in each condition (also when evaluated before the homepage) because of their strong association with the website—either due to the layout (online help) or by the core purpose and popularity of the service (literature search). This in turn means it did not make a difference if the organization "ZBW" was primed because it was already cognitively available. Thus, the order did not matter.

Theoretical and Practical Implications

At a very general level, the findings show that the order in which usability evaluations are conducted for multiple components of a website can matter. Thus, the order of the questions (or scales) should be permuted and possible order effects should be statistically controlled for. In particular, the pattern of findings leads to several more specific insights.

The halo effect demonstrates that the image of the website owner can influence the usability ratings. This is in line with previous findings on the influence of the image of an organization (e.g., university, company), on the licensing rate (Sine, Shane, & Di Gregorio, 2003), or the reputation during crisis management (Coombs & Holladay, 2006). In this context, my findings

show that the image of an organization can also influence the perceived usability of a website with its connected services. In relation to usability evaluation, this means that not only aesthetics and quality of the content can bias usability judgments but also the image of the website owner. In this context, it is worth mentioning that the significant difference found here corresponds with different grades (B versus C). Thus, the bias not only is of theoretical interest but also has practical relevance for user experience.

In addition, the findings demonstrate that such a halo effect by the image of an organization is dependent on the degree of association between the organization (the website owner) and the evaluation object (the online service). From a practical point of view, this implies that different usability ratings of single parts of a website might be due to the degree of association with the website owner and do not necessarily reflect differences in handling. That means that if the association with the website owner is equally strong for all evaluation objects, the evaluations of these objects are systematically biased. However, if the evaluation objects have associations of different degrees, the differences in usability evaluation could be due to the presence or absence of a halo effect. Consequently, "pure" usability ratings require a kind of blind evaluation by the user, for example, by making the logo invisible and using a no-name label as a placeholder. However, it is an open question whether and how this is manageable in practice. Additionally, it has to be taken into account that the aesthetics can influence the ratings. Maybe it is more reasonable to include some kind of control questions—on aesthetics, on the quality of content, and on the image (trustworthiness etc.) of the website owner—that allow for the statistical control of these influences.

The study provided no statistical support for a part-whole effect in relation to the homepage and the associated online services. Even though the homepage represents the website owner, there is no evidence that the homepage is seen as a representative of the entire website. From a practical point of view, this implies that the evaluation of the homepage is not an adequate indicator for the usability of the entire website with its different components.

Limitations of the Study and Open Questions for Future Research

So far, my findings have provided only initial evidence on possible order effects for the evaluation of a website with its single subsections. However, my sample was rather small and the use case was rather specific. Additionally, the prior knowledge of participants might have biased the findings. Even though there was no systematic influence of the assessed control variables (including prior experiences with the ZBW and its services, frequency of PC use and Internet use), these control variables were rather general and did not allow me to investigate the influence of prior knowledge in a systematic way. Also, the additional control analyses on the specific subsample who knew the publishing services provided only preliminary evidence that the halo effect is independent of the users' expertise. For future research, it would be interesting to compare experts versus novices as Lah and Lewis (2016) did in their study. Similarly, it is an open question whether the halo effect can be replicated with unbiased participants. In the reported study, the given ratings were probably based on the participants' prior experience with the homepage and the services. Thus, it would be interesting to test if the same pattern of findings will appear for unbiased participants who use (e.g., in a usability test) the homepage and the services for the first time immediately before they have to fill out the questionnaire.

Additionally, the varying degrees of association with the website owner were not systematically manipulated or measured but merely determined by heuristics. Similarly, there was no systematic experimental variation regarding the website owner's image. To provide more solid evidence on the influence of the image of a website owner, it would be advantageous to have laboratory tests using identical (fictional) websites (i.e., with identical content and aesthetics), while systematically varying the degree of association (strong versus weak) and the image of the website owner (good versus poor). Varying the image of the website owner would also provide insights on whether there is a reverse halo effect, that is, a poor image will lower the usability ratings.

Another limitation of my study relates to the absolute value of the SUS scores. As mentioned above, the difference found here is of practical relevance. However, it has to be taken into account that all the usability scores in the study were rather high, especially the ratings for the homepage that represents the website owner. Thus, it is an interesting open question for future

research if there might be a reverse halo effect in the case of a low usability rating for the homepage. In this context, it might also be interesting to compare the factor structure of the SUS scores for the separate sections of a website to see if the user ratings have the same meaning for all subsections. Because my sample size was too low to perform this analysis, this open question might be addressed by future research.

Conclusion

Despite the limitations of my study mentioned above, the findings demonstrate that the order of the separated evaluation of the single parts of a website can influence the evaluation results under specific conditions. There was not a general halo effect in the sense that a preceding evaluation of the homepage necessarily affects the subsequent ratings of the services. Rather, the halo effect appears to be dependent on the strengths of the association between the website owner and the services. For an obvious association, there will be probably no halo effect. However, for weakly associated services, the order of evaluation can matter. If the association between the website owner and the service is not obvious, the preceding evaluation of the homepage can act as priming of the (good or bad) image of the website owner which in turn will probably create a halo effect that influences the subsequent ratings. Thus, in case of weak (or unclear) associations between the website owner and the components of a website, it is beneficial to test different orders of presentation. This is important in order to prevent a systematic bias of the usability evaluation. However, if the association with the homepage (and the website owner) is obvious for all components, it is rather unlikely that the order of presentation will influence the evaluation results.

As a very general recommendation, practitioners should bear in mind that the perceived usability and the assessed usability ratings can be biased by other influences. Thus, a controlled setting and additional questions on possible influences (aesthetics, quality of content, prior experience, and image of the organization) are beneficial for adequate usability monitoring.

Tips for Usability Practitioners

The order effect found here leads to the following recommendations for practice:

- If the evaluation of a website relates separately to the homepage itself as well as to the other components of the website, the order of the presentation of the pages can influence the evaluation results. Thus, it is beneficial if the order of presentation (of the usability evaluation of the components) is systematically permuted or statistically controlled.
- In situations when there are some orders of presentations that don't make sense or would have an extremely low likelihood, it seems reasonable to provide only the most likely orders because they represent the most realistic scenarios. If more than one order of presentation makes sense, a selective comparison of different orders can provide at least preliminary insights if order matters.
- The need to evaluate different orders of presentation probably depends on the strengths of the association between the components and the website owner. If the association is (partly) rather weak, there is the need for different orders of presentation. If the association with the website owner is obvious for every component, then a halo effect is unlikely and it might be okay to test only one order.
- The image of the website owner can cause a halo effect on the usability ratings. Thus, for a more neutral usability evaluation (especially for rather new unknown products), it could be beneficial to hide, when possible, the connection with the organization (e.g., remove the logo).
- It could be advantageous to include, when possible, participants who have a wide range of experience. Thereby, it is beneficial to assess the prior experiences of the users in a systematic way and use it as control variable.
- Even if the connection with the website owner is not immediately visible for an online service, the association with the website owner can be primed by a prior evaluation of the homepage. Thus, for the evaluation of multiple components of a website, the association with the website owner should be kept constant or the order of the multiple evaluations should be controlled for.

Acknowledgements

Thanks to all people who participated in the study for their patience and cooperation. A special thanks to my colleagues Anika Ostermaier-Grabow and Elisabeth Steinhagen for their help and assistance during this study. Thanks to Klaus Tochtermann, the director of the ZBW, who initiated usability research at the ZBW.

References

- Ahammer, I. M. (1971). Desirability judgments as a function of item content, instruction set, and sex: A life-span developmental study. *Human Development*, 14, 195–207.
- Aljukhadar, M., & Senecal, S. (2015). Determinants of an organization's website ease of use: The moderating role of product tangibility. *Journal of Organizational Computing and Electronic Commerce*, 25, 337–359.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, 89–195). New York: Academic Press.
- Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS score mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4(3), 114–123.
- Bargh, J., & Pietromonaco, P. (1982). Automatic information processing and social perception: The influence of trait information presented outside of conscious awareness on impression formation. *Journal of Personality and Social Psychology*, 43, 437–449.
- Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural equation modelling. *Sociological Methods & Research*, 16, 78–117.
- Berkman, M. I., & Karahoca, D. (2016). Re-assessing the usability metric for user experience (UMUX) scale. *Journal of Usability Studies*, 11(3), 89–109.
- Borsci, S., Federici, S., Bacci, S., Gnaldi, M., & Bartolucci, F. (2015). Assessing user satisfaction in the era of user experience: Comparison of the SUS, UMUX, and UMUX-LITE as a function of product experience. *International Journal of Human-Computer Interaction*, 31, 484–495.
- Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). London: Taylor & Francis. Retrieved July 2011 from <http://hell.meiert.org/core/pdf/sus.pdf>
- Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. *Proceedings of CHI 1988* (pp. 213–218). Washington, DC: ACM.
- Coombs, W. T., & Holladay, S. J. (2006). Unpacking the halo effect: Reputation and crisis management. *Journal of Communication Management*, 10(2), 123–137.
- Davis, F. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340.
- Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality & Social Psychology*, 24(3), 285–290.
- Edwards, A. L. (1957). *The social desirability in personality research*. New York: Dryden.
- Finstad, K. (2010). The usability metric for user experience. *Interacting with Computers*, 22(5), 323–327.
- Gediga, G., Hamborg, K. C., & Dürtsch, I. (1999). The IsoMetrics usability inventory. An operationalisation of ISO 9241-10 supporting summative and formative evaluation of software systems. *Behaviour and Information Technology*, 18, 151–164.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 41–58). New York: Academic Press.

- Grier, R. A., Bangor, A., Kortum, P., & Peres, S. C. (2013). The System Usability Scale: Beyond standard usability testing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 57(1), 187–191.
- Hartmann, J., Sutcliffe, A., & De Angeli, A. (2008). Towards a theory of user judgement of aesthetics and user interface quality. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 15(4, Article No. 15). New York, NY: ACM.
- Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction*, 19, 319–349.
- Kortum, P., & Johnson, M. (2013). The relationship between levels of user experience with a product and perceived system usability. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 57(1), 197–201.
- Lah, U., & Lewis, J. R. (2016). How expertise affects a digital-rights-management-sharing application's usability. *IEEE Software*, 33(3), 76–82.
- Leuthesser, L., Kohli, C. S., & Harich, K. R. (1995). Brand equity: The halo effect measure. *European Journal of Marketing*, 29(4), 57–66.
- Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7, 57–78.
- Lewis, J. R., Utesch, B. S., & Maher, D. E. (2015). Measuring perceived usability: The SUS, UMUX-LITE, and ALTUsability. *International Journal of Human-Computer Interaction*, 31(8), 496–505.
- Linek, S. B. (2015). Scale on the quality of literature lists (SQuaLL): Quick quality check from the user perspective. *Proceedings of the 9th International Technology, Education and Development Conference (INTED 2015)*. Madrid, 2nd - 4th of March, 2015.
- Linek, S. B., & Tochtermann, K. (2011). Assessment of usability benchmarks: Combining standardized scales with specific questions. *International Journal of Emerging Technologies in Learning (iJET)*, 6(4), 56–64.
- Linek, S. B., & Tochtermann, K. (2015). Paper prototyping: The surplus merit of a multi-method approach. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 16(3), Article 7. Available at <http://www.qualitative-research.net/index.php/fqs/article/view/2236>
- Mason, R., Carlson, J. E., & Tourangeau, R. (1994). Contrast effects and subtraction in part-whole questions. *Public Opinion Quarterly*, 58, 569–578.
- McLellan, S., Muddimer, A., & Peres, S. C. (2012). The effect of experience on System Usability Scale ratings. *Journal of Usability Studies*, 7(2), 56–67.
- Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, 64, 482–488.
- Murphy, L., Stanney, K., & Hancock, P. A. (2003). The effect of affect. The hedonic evaluation of human-computer interaction. *Proceedings of Human Factors and Ergonomics Society, 47th Annual Meeting 2003* (Vol. 47, pp. 764–768). Retrieved March 2016 from <http://pro.sagepub.com/content/47/4/764.full.pdf+html>
- Nunnally, J. C. (1975). Psychometric theory. 25 years ago and now. *Educational Researcher*, 4(10), 7–14 & 19–21.
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Mahwah, NJ: Erlbaum.
- Prümper, J. (1999). Test IT: ISONORM 9241/10. In H.J. Bullinger & J. Ziegler (Eds.), *Human-Computer Interaction - Communication, Cooperation, and Application Design* (pp. 1028–1032). Mahwah, NJ: Lawrence Erlbaum Associates.

- Raita, E., & Oulasvirta, A. (2010). Too good to be bad: The effect of favourable expectations on usability perceptions. *Proceedings of Human Factors and Ergonomics Society, 54th Annual Meeting 2010* (pp. 2206–2210). Retrieved March 2016 from <http://pro.sagepub.com/content/54/26/2206.full.pdf+html>
- Sauro, J., & Lewis, J. R. (2016). *Quantifying the user experience* (2nd ed.). Boston, MA: Morgan Kaufmann.
- Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments in question form, wording and context*. New York: Academic Press.
- Schwarz, N. (1999). How the questions shape the answers. *American Psychologist*, 54(2), 93–105.
- Schwarz, N., Strack, N., & Mai, H.-P. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly*, 55, 3–23.
- Sine, W. D., Shane, S., & Di Gregorio, D. (2003). The halo effect and technology licensing: The influence of institutional prestige on the licensing of university inventions. *Management Science*, 49(4), 478–496.
- Strack, F., Martin, L. L., & Schwarz, N. (1988). Priming and communication: The social determinants of information use in judgments of life satisfaction. *European Journal of Social Psychology*, 5(18), 429–442.
- Sütterlin, B., & Siegrist, M. (2015). Simply adding the word fruit makes sugar healthier: The misleading effect of symbolic information on the perceived healthiness of food. *Appetite*, 95, 252–261.
- Thorndike, E. L. (1920). A constant error in psychological rating. *Journal of Applied Psychology*, 4, 25–29.
- Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers*, 13, 127–145.
- Tuch, A. N., Roth, S. P., Hornbæk, K., Opwis, K., & Bargas-Avila, J. A. (2012). Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI. *Computers in Human Behavior*, 28, 1596–1607.
- Tullis, T. S. & Stetson, J. N. (2004). A comparison of questionnaires for assessing website usability. Retrieved June 2017 from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.396.3677&rep=rep1&type=pdf>

About the Author



Stephanie B. Linek

Dr. Linek is an expert in usability evaluation and media psychology at the ZBW—Leibniz Information Centre for Economics (Kiel, Germany). Her research interests include HCI, social media, usability, evaluation, and methodology.