Linek, Stephanie

Conference Paper

Scale on the quality of literature lists (SQuaLL): Quick quality check from the user perspective


This version is available at:
http://hdl.handle.net/11108/196
SCALE ON THE QUALITY OF LITERATURE LISTS (SQUALL): QUICK QUALITY CHECK FROM THE USER PERSPECTIVE

Stephanie B. Linek
ZBW – Leibniz Information Centre for Economics (GERMANY)

Abstract

Nowadays, a modern digital library or Library 2.0 [1] usually offers an online portal for literature search. Thereby, it is of special importance to assure a good usability as well as an obvious excellent quality of the search portal. For the assessment of usability, several standardized scales exist, e.g., the System Usability Scale (SUS, see [2]) that offers an economic way for a global usability check. However, the assessment of the quality of the listed results of the literature search is more complicated, because the different professional search portals offer very special and partly unique features. The related assessment instruments are often very detailed and focus more on systems of the digital library in general as on the literature search portals in specific.

Beyond this background, this paper describes the newly developed Scale on the Quality of Literature Lists (SQuaLL). The SQuaLL offers a quick and easy way to assess the quality of a literature search portal from the view of the user. The SQuaLL was originally developed with respect to the literature search portal EconBiz (www.econbiz.de), but can be easily modified for the evaluation of other academic literature search services. The sum-score of the SQuaLL is conceptualized as a general quality index that can be used as an additional supplement to a usability evaluation. The general quality index is defined as a composition of ten selected quality criteria. The identified quality criteria are based on an expert query.

The format of the scale is constructed analogous to the SUS. It comprises ten items that regard to ten characteristics that reflect the core quality criteria of academic search portals: Five of these items assess the “must haves”, i.e., the basic characteristics that every literature search portal should offer (e.g., a reasonable ranking and filtering). Another four items regard to the “nice to have”, i.e., features that are part of most modern search portals because they are of special interest for the user (e.g., availability of full text download). One item refers to a rather special feature of EconBiz, i.e., the event search. The wording of this special item can be easily modified with respect to different specifics of other search portals. All ten items are formulated in form of statements that have to be rated on a 5-point Likert scale. The possible sum-scores of the SQuaLL are between 0 (very poor quality) and 100 (very high quality).

The practical use and the parallel test reliability of the SQuaLL were already proven for a controlled offline sample. For the analysis of the parallel test reliability two forms (A and B) of the SQuaLL were established. The correlation between form A and form B of \( r = .777 \) (p < .001) indicates a good parallel test reliability of the sum-scores. Additionally, also an online version of the SQuaLL was successfully applied.

To sum up, the SQuaLL offers a global quality index of literature search portals from the user perspective. The items of the scale can be easily modified for the evaluation of different academic search portals of multiple scientific domains. The practical and theoretical implications will be discussed.

Keywords: digital library, Library 2.0, search portal, evaluation, quality check, user perspective, information management.

1 INTRODUCTION

The core service of a modern digital library or Library 2.0 [1] is usually the portal for professional literature search. But nowadays many students and researchers rely on the search by Google, because of the easy handling and the seemingly sufficient quality – even though they are aware of the shortcomings of the Google search. Thus, it is of special importance for a scientific Library 2.0 to assure a good usability as well as an excellent and obvious quality of the professional search portals.
For the assessment of usability there exist several standardized scales, e.g., the System Usability Scale (SUS, see [2]). The SUS is a short scale with ten items that offers a very economic way of usability evaluation. Another big advantage of the SUS is the general formulation of the items which enables the application on very diverse software products or Internet sites (including search portals).

Besides the usability of academic literature search portals, also the quality of the listed results of the literature search (hereinafter called “literature lists”) is of essential importance. Thereby, the assessment of the quality of literature lists is often complicated because professional search portals of different digital libraries have often very special and partly unique features (contrariwise to the simple Google search). Additionally, it has to be taken into account that the overall quality of literature lists comprises on the one hand aspects of the perceived usefulness and on the other hand aspects of scientific value. Based on these considerations the Scale on the Quality of Literature Lists (SQuaLL) was developed. The SQuaLL assesses the usefulness as well as the perceived scientific value from the user perspective. It was designed analogous to the SUS as a short scale for a quick and easy evaluation. For a better understanding of the scientific background, in the following I provide a short glance on usability and usefulness before I explore the general aim of the SQuaLL.

1.1 Usability and Usefulness

Nowadays usability is a big issue. However, the differentiation between usability and usefulness is partly confusing. In literature there are several slightly different definitions of usability. Besides accessibility, most definitions refer to four aspects: effectiveness, efficiency, appropriateness for subjective aims of the user, and joy of use (e.g., [3], [4]). The usefulness (or utility) is the perceived ability of a certain content or functionality to provide support for a challenge (or problem) of the user. Usefulness can be assumed as (at least partly) included in very broad definitions of usability in the sense of “appropriateness for subjective aims”. However, other conceptions differentiate usability of handling from the usefulness of the application. For example the Technology Acceptance Model (TAM) of [5] or the P3-Model of [6] make an explicit distinction between usability versus usefulness (utility). Thereby, the relation between usability and usefulness is conceptualized as pivotal factor for the user acceptance and future use.

There are also more specific approaches on the usefulness of digital libraries [7] [8] [9] [10]. Generally spoken, these approaches conceptualized digital libraries and bibliographic database in the form of specific components. Thereby, usefulness of the system is understood as the relation between the user and the other components (content, functionalities). The accordingly measurement instruments (e.g., [9]) focus on the abstract criteria of these concepts in a rather general way. This enables to apply the questionnaire on many various systems. However, due to the very abstract formulations the outcomes can be hardly connected to the concrete practical features. Additionally, the according scales assess rather the total system around the digital library’s usage and less the concrete quality (in form of usefulness and scientific value) of the online search portal.

Beyond this background the SQuaLL was established as an economic and quick assessment instrument for the quality of academic literature search portals from the user perspective. Thereby, quality was conceptualized as combination of usefulness and scientific value.

1.2 Aim of the SQuaLL

When evaluating a professional search portal, it is important to note, that users and information professionals (like librarians) have different requirements. Accordingly, [8] stated that for an expert-based evaluation (by librarians) versus a user-based evaluation other evaluation criteria are needed. However, this is only half of the story. A pivotal aspect that is often neglected is the question if and to which extent the users appreciate the functionalities that are estimated as core characteristics by librarians or information managers. The answer to this question provides not only feedback on the usefulness but also delivers a quality judgment from the users’ view.

In this line of reasoning, the SQuaLL was developed for the assessment of a general global quality index of professional literature search portals. This quality assessment is conceptualized as complement of the usability evaluation. Thereby, the quality index of the SQuaLL aimed not only at the usefulness but also regards to the scientific and professional value of the listed search results.

The items of the SQuaLL regard to ten core characteristics of academic literature search portals. The wording of the items is formulated in a way that the usefulness and the quality can be judged from the users’ view. This rational enables an overall quality judgment from the user perspective with respect to
those characteristics that are assumed essentially by the librarians. It’s important to note, that the assessed characteristics are not abstract but rather very close to the concrete practical functionalities and outputs of the literature search portal. Another crucial point of the SQuaLL is its flexibility. The items are formulated in analogy to the SUS, i.e., in a way that the item can be easily adapted to different features of diverse literature search portals.

2 DEVELOPMENT OF THE SQUALL

2.1 Practical Background of the SQuaLL

Background of the construction of the SQuaLL was a very practical one: an upcoming complete relaunch of the website of the world largest library for economics, namely the ZBW (www.zbw.eu). As a modern Library 2.0, the ZBW has several online services. The usability of the ZBW’s online services is continuously monitored and systematically improved. (A general description of the usability evaluation of a Library 2.0 can be found by [11]). One essential cornerstone in this iterative usability evaluation was the measurement of the usability baseline before the relaunch. This baseline serves as benchmark for comparison with the usability of the new website and the upcoming improvements (for details please see [12]). The usability evaluation was mainly done by means of the System Usability Scale (SUS, see [2]) and the ISONORM [13]. Accordingly, a benchmarking questionnaire was developed, that assessed the usability of the homepage as well as of the three main online services including the literature search portal EconBiz (www.econbiz.eu). For this core service, the product managers of EconBiz wanted also information beyond usability, namely on the general user evaluation of the contents and the general quality of EconBiz. Therefore, a special scale was constructed: The Scale for the Quality of the Literature List (SQuaLL).

2.2 Construction of the SQuaLL

2.2.1 Identification of Quality Criteria

As explained above, the SQuaLL was designed for the assessment of the general quality of the search results and the listed literature, respectively. The general quality was defined by a composition of selected quality criteria with respect to usefulness and scientific value. The selection of these specific quality criteria was done in close cooperation with the internal experts of the search portal EconBiz. Therefore an expert query in subsequent steps was made. The first step was a semi-structured interview with three internal experts. Based on these qualitative data, a list of quality criteria and accordingly items was formulated. This list was given to the former interrogated experts as well as to a small group of six associated experts that were not interviewed before. They were asked to give comments on the priority and importance of the listed quality criteria as well as on the validity (fitting accuracy) of the wording of the associated items.

Based on this expert query, the following quality criteria were identified:

- Five core criteria, that can be seen as “must haves” of every literature search portal, namely exhaustiveness, a reasonable ranking, scientific proofed quality, relevance (based on metadata), as well as valid and modifiable filters.
- Four additional features that nowadays are integrated in most professional literature search portals, because they are of special interest for the user. These four “nice to have” features are: availability of full text is visible, export functionality, additional (helpful) information, and information on new releases is available
- One special feature of EconBiz that is beyond the general standard of literature search portals, namely the event search that included also important events (related to the keywords) in the search.

Each of these ten quality criteria is addressed by one item of the SQuaLL.

2.2.2 Format of the SQuaLL

The items and the answering format of the SQuaLL is designed analogous to the SUS, i.e., it comprises ten statements that have to be rated on a 5-point Likert Scale. However, contrariwise to the SUS there are no inverted items included, because of the variety of the identified quality criteria.
Each of the ten items refers to one of the identified quality criteria listed above: Five of the items assess the “must haves”, i.e., they relate to the basic core characteristics that every literature search portal should offer (e.g., a reasonable ranking and filtering). Another four items regard to the “nice to have”, i.e., features that are part of most modern search portals because they are of special importance for the user (e.g., availability of full text download). One item refers to a rather special feature of EconBiz, namely the event search. The special item is formulated as such that it can be easily modified with respect to different specifics of other search portals, e.g., instead of “event search” one could ask for “multimedia search”.

The global quality index of the SQuaLL is conceptualized as a sum-score calculated analogous to the SUS. The possible values are between 0 and 100. Details on the calculation of the sum-score are given in the chapter on the description of the SQuaLL.

2.2.3 Pilot Test

A first draft of the scale 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 assure the comprehensibility and appropriateness of the wording. Based on this pilot test, some slight changes in the wording of questions were made.

2.2.4 Two Parallel Versions of the SQuaLL (Form A and Form B)

Since the scale was newly constructed and thus reliability was not proven so far, I formulated and applied two parallel versions, form A and form B. The form B of the SQuaLL assesses the identical quality criteria like the original form A of the SQuaLL. In contrast to form A the wording of the items of form B is slightly modified and the order of the corresponding items of the two forms is intermixed. Additionally, form B contains one item that has an explicit example in the wording. This latter modification was also meant as an exploratory pilot test if and how such examples in brackets might influence the answers.

2.2.5 Online Version

About one year later also an online version of the SQuaLL was established. For the online version of the SQuaLL the item on the event search was slightly modified, because of some changes of the search portal EconBiz (the results of the event search were separated from the literature results list). Details are given in the chapter about the online version.

3 DESCRIPTION OF THE SQUALL AND ITS APPLICATION

Like already described, each of the identified ten quality criteria is assessed by one item of the SQuaLL. Each of the items has to be rated on a 5-point Likert scale from 1 = totally disagree to 5 = totally agree. Like mentioned above, the item-scores and the sum-scores of the SQuaLL are calculated analogously to the formula of the SUS (see [2]).

The score of each item of the SQuaLL is calculated by the following formula:

\[
\text{item-score} = \text{position} - 1
\]

The overall sum-score of the SQuaLL is calculated by the following formula:

\[
\text{sum-score SQuaLL} = \Sigma (\text{item-score}) \times 2.5
\]

The possible values of the SQuaLL sum-score are between 0 and 100. Thereby, 0 indicate a very poor quality and 100 reflect an excellent quality.

An overview on the assessed quality criteria and the accordingly items of the SQuaLL (form A and form B) is given in table 1. The original scale (form A and form B) is in German. The given English translation for the items is very close to the German wording (even if it is no good English style) to underline the modifications in wording. Interested readers can contact the author for the German original scale.
Table 1: Quality criteria and wording of the items of the SQuaLL (form A and form B)

<table>
<thead>
<tr>
<th>Quality Criteria</th>
<th>Form A</th>
<th>Form B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nr.</td>
<td>A-item</td>
</tr>
<tr>
<td>Exhaustiveness</td>
<td>1</td>
<td>The result list of EconBiz provides me an exhaustive overview on the subject area.</td>
</tr>
<tr>
<td>Reasonable ranking</td>
<td>2</td>
<td>The ranking of the result list of EconBiz is well elaborated.</td>
</tr>
<tr>
<td>Scientific proofed quality</td>
<td>3</td>
<td>I can be sure, that the listed contents of EconBiz are of high scientific quality.</td>
</tr>
<tr>
<td>Relevance (based on meta data)</td>
<td>4</td>
<td>The listed results in EconBiz are in very good accordance with the used search keys.</td>
</tr>
<tr>
<td>Valid and modifiable filters</td>
<td>5</td>
<td>The filter functionalities of EconBiz are flexible modifiable and select relevant items form irrelevant issues in a reliable way.</td>
</tr>
<tr>
<td>Events included</td>
<td>6</td>
<td>I estimate the listing of events within the literature list as very informative and interesting.</td>
</tr>
<tr>
<td>Availability of full texts is visible</td>
<td>7</td>
<td>From the result list of EconBiz I can immediately recognize, if and where a full text of the reference is available.</td>
</tr>
<tr>
<td>Export functionality</td>
<td>8</td>
<td>I can export the listed references without any problems, e.g., integrate it in my own literature list or software tools for managing references.</td>
</tr>
<tr>
<td>Additional information¹</td>
<td>9</td>
<td>For the listed references of the result list of EconBiz there is helpful additional information given.</td>
</tr>
<tr>
<td>Information on new releases available</td>
<td>10</td>
<td>Via EconBiz I can also look for new releases.</td>
</tr>
</tbody>
</table>

¹ This was the only item with an explicit example in the wording in form B (but not in form A).

* In the online version of the SQuaLL the item on the event search was slightly modified; details are given in the chapter about the online version.
PARALLELTEST RELIABILITY AND PRACTICABILITY OF THE SQUALL

4.1 Background of the Study

The study on the paralleltest reliability of the SQuaLL was integrated in a big survey for the overall evaluation of the ZBW's homepage and the integrated three main online services (see section 2.1 on the practical background of the SQuaLL). As indicator for the paralleltest reliability, the correlations between the sum-scores as well as the correlations between the corresponding parallel items of form A and form B were calculated.

4.2 Methodology

The SQuaLL (form A) was embedded as additional scale in a long benchmarking questionnaire. Like already explained this benchmarking questionnaire aimed at a first baseline measurement of the usability of the ZBW's website (see section 2.1). Besides the SQuaLL the benchmarking questionnaire comprised the SUS (separate scales for the homepage, and the three services) and the ISONORM for a detailed evaluation of EconBiz. Furthermore, some open questions on the ZBW, a scribbling task, and several control variables were included (age, gender, occupation, computer and Internet experience). For a detailed description of the benchmarking questionnaire please see [12]. Form A of the SQuaLL was presented after the standardized scales for usability evaluation (SUS, ISONORM) and some general open questions, but before the qualitative scribbling task. Form B was presented at the very end of the study, i.e., after the benchmarking questionnaire was completed and the participant had received a short face-to-face post-interview. When answering form B of the SQuaLL, the form A was not available to the participant.

4.3 Procedure

The prerequisites for the participation in the study were that the participants knew and used the ZBW homepage as well as the service EconBiz before. It was not an obligation to know also the other two services (EconDesk and EconStor). These prerequisites for participation mirrored the majority of the usual ZBW clients. The participants were recruited personally in the buildings (reading room) of the ZBW at the locations Hamburg and Kiel. Additionally, flyers and mouth-to-mouth recruitment were used.

Altogether 96 participants completed the benchmarking questionnaire. As incentive each participant received a 20,- € voucher for a popular online shop. During the completion of the questionnaire, some sweets were offered. The duration of the test sessions varied between a half and one and a half hour. The questionnaire was answered in group sessions in a separate room in the ZBW buildings in Hamburg or in Kiel. After a short welcome, the participants were informed that their data were threatened anonymously and strictly confidential. The participants were instructed (orally and in the written instructions) to work on the questions in the given order. During the whole study an instructor was present. The long questionnaire with form A of the SQuaLL (in the middle) was handled out and answered in a written paper-pencil based format. After the completion of the questionnaire, the instructor took the completed materials and made a short post-interview with each participant. Subsequently, the participants received the form B of the SQuaLL on a separated sheet of paper. (Please note: At this point in time, the form A of the SQuaLL was not available to the participant.) After the participants had filled out the form B they received the voucher as their reward for the participation and had again the opportunity to ask questions or make comments. The instructor requested them not to talk about the study for the next few weeks.

4.4 Results

4.4.1 Description of the Sample

The sample contains 47 male and 49 female participants (together 96 persons). The age of the participants was at average 27.33 years and varies between 19 and 53 years. Most of the participants were students (80%) of different fields of economics. The majority used the PC (89.6 %) as well as the Internet (91.7%) daily for several hours. The frequency and duration of information search via the Internet was a bit lower. Their self-rated competence (on a 10-point rating scale) with PC (m = 7.66) and Internet (m = 8.18) was relatively high. The self-rated competence with information search via the Internet was on a comparable level (m = 7.33). At average the participants use the ZBW (official building) since 3.12 years and the Website of the ZBW (online use) since 2.61 years. The physical
library, i.e. the building of the ZBW was visited of half of the participants less than a week. The online visits were more frequently. All participants (n = 96) were familiar with the ZBW’s homepage and the literature search service EconBiz.

4.4.2 Descriptive Statistics and Comparison of Means

Overall, the literature search portal EconBiz received medium to high ratings for both parallel versions of the SQuaLL. In the according t-tests for dependent groups no significant differences between the means (form A versus form B) were found. The descriptive statistics of the sum-scores and item-scores of the form A and form B can be found in the table 2 below.

Table 2: SQuaLL – form A and B: means (standard deviation in brackets)

<table>
<thead>
<tr>
<th>Quality criteria</th>
<th>SQuaLL A</th>
<th>SQuaLL B</th>
</tr>
</thead>
<tbody>
<tr>
<td>sum-score</td>
<td>62.29 (11.74)</td>
<td>66.07 (10.74)</td>
</tr>
<tr>
<td>Exhaustiveness</td>
<td>item A-1 2.38 (0.94)</td>
<td>item B-5 2.58 (0.93)</td>
</tr>
<tr>
<td>Reasonable ranking</td>
<td>item A-2 2.24 (0.79)</td>
<td>item B-8 2.54 (0.89)</td>
</tr>
<tr>
<td>Scientific proofed quality</td>
<td>item A-3 3.20 (0.71)</td>
<td>item B-10 3.31 (0.69)</td>
</tr>
<tr>
<td>Relevance</td>
<td>item A-4 2.71 (0.82)</td>
<td>item B-2 2.56 (0.79)</td>
</tr>
<tr>
<td>Valid and modifiable filters</td>
<td>item A-5 2.55 (0.90)</td>
<td>item B-7 3.00 (0.82)</td>
</tr>
<tr>
<td>Events included</td>
<td>item A-6 1.99 (1.05)</td>
<td>item B-3 2.13 (0.95)</td>
</tr>
<tr>
<td>Availability of full texts is visible</td>
<td>item A-7 2.72 (1.07)</td>
<td>item B-9 3.08 (0.83)</td>
</tr>
<tr>
<td>Export functionality</td>
<td>item A-8 2.17 (0.94)</td>
<td>item B-6 2.27 (0.81)</td>
</tr>
<tr>
<td>Additional information</td>
<td>item A-9 2.55 (0.74)</td>
<td>item B-1 2.22 (1.03)</td>
</tr>
<tr>
<td>Information on new releases</td>
<td>item A-10 2.43 (0.95)</td>
<td>item B-4 2.73 (0.89)</td>
</tr>
</tbody>
</table>

4.4.3 Correlative Analyses: Paralleltest Reliability

The correlative analyses for the association between form A and form B were done for the sum-score as well as for the ratings of the single items. The correlation of the sum-scores of form A and form B was significant with a value of $r = .777$. This indicated a good paralleltest reliability for the SQuaLL sum-score as a global quality-rating. The correlations of the parallel items varied between $r = .698$ (for “scientific proof quality”) and $r = .144$ (for “additional information”). Detailed statistical information on the correlations is listed in table 3.

Table 3: SQuaLL: Pearson correlation between sum-score and the single items of form A and form B, significance (p) and number of valid cases (n). Significant correlations are marked with *

<table>
<thead>
<tr>
<th>Quality criteria</th>
<th>SQuaLL A</th>
<th>SQuaLL B</th>
<th>Correlation (Pearson)</th>
<th>p</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>sum-score</td>
<td>sum-score B</td>
<td>.777*</td>
<td>.000</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>Exhaustiveness</td>
<td>item A-1</td>
<td>item B-5</td>
<td>.442*</td>
<td>.000</td>
<td>95</td>
</tr>
<tr>
<td>Reasonable ranking</td>
<td>item A-2</td>
<td>item B-8</td>
<td>.574*</td>
<td>.000</td>
<td>96</td>
</tr>
<tr>
<td>Scientific proofed quality</td>
<td>item A-3</td>
<td>item B-10</td>
<td>.698*</td>
<td>.000</td>
<td>96</td>
</tr>
<tr>
<td>Relevance</td>
<td>item A-4</td>
<td>item B-2</td>
<td>.417*</td>
<td>.000</td>
<td>96</td>
</tr>
<tr>
<td>Valid and modifiable filters</td>
<td>item A-5</td>
<td>item B-7</td>
<td>.468*</td>
<td>.000</td>
<td>96</td>
</tr>
<tr>
<td>Events included</td>
<td>item A-6</td>
<td>item B-3</td>
<td>.474*</td>
<td>.000</td>
<td>96</td>
</tr>
<tr>
<td>Availability of full texts is visible</td>
<td>item A-7</td>
<td>item B-9</td>
<td>.594*</td>
<td>.000</td>
<td>96</td>
</tr>
<tr>
<td>Export functionality</td>
<td>item A-8</td>
<td>item B-6</td>
<td>.630*</td>
<td>.000</td>
<td>96</td>
</tr>
<tr>
<td>Additional information¹</td>
<td>item A-9</td>
<td>item B-1</td>
<td>.144</td>
<td>.160</td>
<td>96</td>
</tr>
<tr>
<td>Information on new releases</td>
<td>item A-10</td>
<td>item B-4</td>
<td>.488*</td>
<td>.000</td>
<td>96</td>
</tr>
</tbody>
</table>

¹ The only item with an explicit example (“download-statistics”) in form B but not in form A
All correlations were significant with one remarkable exception, namely the correlation between the parallel items on “additional information”. It is worth noting, that only for the items on “additional information” there was an explicit example in the wording in form B but not in form A. None of the other items included explicit examples.

To sum up the findings, the correlation between the sum-scores of the parallel forms of the SQuaLL indicates a good paralleltest reliability. However, the paralleltest reliability of most of the single items was not sufficiently high enough to use the item-scores for detailed insights on single quality criteria. Only the item on “scientific proof quality” ($r = .698$) might be assumed as reliable for an isolated interpretation.

4.5 Resume

The data suggest a good paralleltest reliability of the sum-score of the SQuaLL. This finding provides first evidence that the SQuaLL sum-score can be used as a reliable global quality index. However, most of the correlations between the single parallel items are comparably low (even though significant). Thus, the single items are not apt for an isolated (reliable) interpretation. Only the items on scientific proved quality show a sufficiently high paralleltest reliability and might be used as an isolated quality criteria for specific interpretations.

Another interesting finding is the low correlation between the two parallel items on “additional information”. For the assessment of this criterion an example (in brackets) was included in the wording of form B but not in form A. It’s a remarkable coincidence, that the only item with an example was also the only item that showed not a significant correlation with the corresponding parallel item. This finding suggests that concrete examples might distort the ratings of the users. One possible reason could be that the mentioned example was overestimated in the rating. Thus, the mention of specific examples should be avoided.

Overall, the SQuaLL shows a good paralleltest reliability as a global quality index, but the values of most of the single items – except the item on scientific proofed quality – should not be used for specific interpretations.

5 ONLINE APPLICATION OF THE SQUALL

5.1 Background and Methodology

About one year after the described paper-pencil application of the offline benchmarking questionnaire (including the SQuaLL), an online version was established. The online benchmarking questionnaire was a shortened version of the according offline version. The shorter online version of the benchmarking questionnaire aimed at a repeated and less resource intensive online evaluation of the ZBW’s website and its services. As standardized scales it comprised only the SUS for the ZBW’s homepage and the three services as well as the SQuaLL (form A). Additionally, a shortened set of control variables and qualitative questions on the ZBW were integrated.

During the one-year-break there were no substantial changes of the ZBW’s Website, neither the homepage nor the services, namely EconBiz. The only small modification of EconBiz was the separation of the event search from the results list of the literature references. Thus, the wording of the item on the event search was slightly modified (see table 1 in section 3).

The implementation of the online survey was done by the software SoSci Survey (https://www.soscisurvey.de/?l=eng). After the implementation of the questionnaire, internal technical tests for correct filters, typos etc. were made. Additionally, checks for technical errors and understandability of the modified textual parts were made.

The link to the online questionnaire was sent out by email or was embedded in an according announcement at the homepage and the Facebook-page of the ZBW. There was no back-button in the online questionnaire, i.e., the participants couldn’t change their answers after they moved to the next question. For the rating-scales there were forced answers, i.e., only if all items were answered, the participants could go forward to the next question. All data were assessed in an anonymous way. Prerequisities for participation (analogous to the offline version) were that the participants knew and used the ZBW Homepage as well as the service EconBiz before.
Since there were only minor changes of the ZBW’s website and EconBiz, the related descriptive quantitative evaluation results (i.e., means) should be on an equal level like the offline evaluation reported above. However, this is not a methodological controlled retest reliability test. Methodological restrictions of this comparison are due to the following reasons:

- One modified functionality of EconBiz and slightly changed wording of the according item of the SQuaLL
- Different modes of assessment (offline versus online; shortened version of the benchmarking questionnaire)
- Different ways of recruitment and an accordingly different sample of participants

Nevertheless, the comparison between the offline version versus the online version of the SQuaLL can be seen as a rough indicator for the comparability and equivalency of the online version and the offline version. Details on the comparison of the offline version versus the online version can be found in [14].

5.2 Results

The online sample consisted of 73 persons (45 male and 28 female). Their age was at average $m = 29.48$ years. Most of them were students (34.2%), employed (28.8%) or both (30.1%). All participants were familiar with the ZBW’s homepage and EconBiz (because this was the prerequisite for participation).

The practical online application of the SQuaLL didn’t show any problems.

As an estimation of the comparability of the offline versus online application of the SQuaLL, t-tests for independent groups were calculated. The non-significant findings of the t-tests indicate that the online version showed comparable results to the offline version. (For details on the comparability of the offline versus online assessment see [14]).

6 RESUMEE AND OUTLOOK

In this paper a newly designed Scale on the Quality of Literature List (SQuaLL) was described. The SQuaLL aimed at a global quality evaluation for literature search portals of modern Libraries 2.0 and can be used as supplement to an usability evaluation.

The newly constructed SQuaLL was empirically tested in an offline setting by means of two parallel versions (form A and form B). The sum-score of the SQuaLL shows good parallel test reliability. However, the single items should not be used for isolated evaluations except the item on “scientific proofed quality”. The SQuaLL was also successfully applied in an offline version (paper-pencil) as well as in an online version (Internet-survey). Both versions deliver equivalent evaluation results.

Overall, the reported findings indicate that the SQuaLL can be used for a reliable quality evaluation of literature search portals. The global quality index of the SQuaLL is complementary to indices of usability evaluation.

It’s worth mentioning, that the SQuaLL can be easily modified and adapted to divers literature search portals of different scientific domains. This is not only important because different search portals offer different functionalities. Rather, it has to be taken into account that the development and improvement of literature search portals is an ongoing process. In the course of technical developments and evolving requirements of the users, also the appropriateness of the selected quality criteria might change. Even though the five core quality criteria (“must haves”) seems to be basic essentials also in the future, the appropriateness of the other criteria (“nice to haves” and special features) might change. The quality of search portals is not a fixed value but rather a flexible concept that has to be defined with respect to the ongoing development of technology and user expectations.

REFERENCES


