Article

Journal of Librarianship and Information Science
2021, Vol. 53(1) 130–152

DOI: 10.1177/0961000620935505
journals.sagepub.com/home/lis

Digital library evaluation measures in academic settings: Perspectives from scholars and practitioners

Iris Xie
School of Information Studies, University of Wisconsin–Milwaukee, USA

Soohyung Joo
School of Information Science, University of Kentucky, USA

Krystyna K. Matusiak
Research Methods and Information Science Department, University of Denver, USA

Abstract

Evaluation criteria and appropriate measures are critical to the success of digital library evaluation. The key problem is a lack of specific measures, especially from a heuristic perspective, corresponding to diverse dimensions and criteria. This study explored a variety of measures for digital library evaluation focusing on their appropriateness. It also compared the similarities and differences in perceptions of the appropriateness of digital library evaluation measures of two groups. Sixty-one participants were recruited representing scholars and academic digital librarians. The participants were instructed to fill in an in-depth survey consisting of 174 measures associated with 10 evaluation dimensions and 82 criteria. The findings of this study highlight the most appropriate measures in each dimension and show significant agreement in identifying appropriate measures by digital library scholars and librarians. Differences in the rating of the measures were found between the two groups in the dimensions of "interface design," "system and technology," "effects on users," "administration," "user engagement," and "context." The significance of this study lies not only in the integration of the perspectives on measures from both researchers and practitioners, but also in providing an inclusive list of measures to guide practitioners to effectively evaluate digital libraries in academic settings.

Keywords

Evaluation measures, evaluation criteria, digital libraries, scholars, digital librarians

Introduction

Digital libraries are defined as representations of emergent and complex forms of digital information collection, organization, design, storage, retrieval, and dissemination at various stages of development. The definition of a digital library is a dynamic concept that is constantly evolving (Saracevic, 2004). The evolution of digital libraries has been marked by three major phases. First, during the formative years of 1991–2001, visionary and experimental projects were translated into real digital library practice. Digital library initiatives and early digital library projects have played a critical role in the development of digital library architecture, standards, iterative design, usability studies, and best practices. Second, when building content and open access in the 2000s, massive digitization projects and open-access repositories represented two characteristics of digital library development. While massive digitization projects enhanced digital library technology, the creation of open-access repositories raised issues concerning the complexity and diversity of digital library development. Lastly, starting in 2010, large-scale digital libraries, such as the Digital Public Library of America, HathiTrust, and Europeana Collections, are signified by their large

Corresponding author:
Iris Xie, School of Information Studies, University of Wisconsin–Milwaukee, Northwest Quad Building B, 2025 E Newport, Milwaukee, WI 53211, USA.
Email: hiris@uwm.edu

collection size, diversity of formats, general and specific collection development policies, level of access, and interoperability (Xie and Matusiak, 2016).

Digital library evaluation is crucial to digital library development and enhancement. Without evaluation, the success of a digital library cannot be valued, the problems of a digital library cannot be identified, and, most importantly, the life cycle of digital library development cannot be sustained. Digital library evaluation consists of the following main components: evaluation dimensions, criteria, and measures (Xie and Matusiak, 2016). These components determine what to assess and how to assess it. The identification of these components provides guidelines for digital library researchers and practitioners to assess different types of digital libraries. Digital library evaluation research has focused on the development of digital library evaluation models and frameworks, as well as the assessment of actual digital libraries. On the one hand, researchers have tried to identify the main dimensions and constructs of digital library evaluation: content; system, technology, and functionalities; interface; users; process and interaction; services; and contexts (Gonçalves et al., 2004; Saracevic, 2004; Xie, 2006). On the other hand, less research has been devoted to the measures corresponding to each of the digital library evaluation criteria.

Few studies have investigated digital library measures from a heuristic point of view. Most have only touched on measures as part of a digital library evaluation model or framework (Fuhr et al., 2001; Zhang, 2010). The majority of articles on measures are related to usability and accessibility studies of digital libraries (e.g. Inal, 2018; Jeng, 2005). The existing research and practice set a good foundation for this study. Nevertheless, inadequate research has been conducted on digital library measures in terms of what are the most and least appropriate measures for the corresponding digital library evaluation criteria, especially from the perspectives of two main stakeholders: digital librarians and digital library scholars. In this study, digital librarians refer to librarians who are mainly responsible for the development and management of digital libraries; digital library scholars refer to researchers who conduct studies related to different aspects of digital libraries. To avoid repetition, the terms "librarians" and "scholars" are used to represent the two groups of digital library stakeholders in this article. This study intends to holistically examine the appropriateness of evaluation measures of multiple dimensions of digital libraries from the viewpoints of library practitioners and scholars. The findings of this study contribute to the field by (1) offering librarians comprehensive and highly appropriate measures to assess digital libraries; (2) identifying the gaps for more research on the measures that are perceived as less appropriate for different types of digital library evaluation dimensions/criteria; and (3) ensuring the validity of these measures by comparing and integrating the perspectives of both librarians and scholars.

Literature review

Digital library evaluation frameworks/models

There are several theoretical models and frameworks that can be used for digital library evaluation. As an early effort, the DELOS Digital Library Reference Model has served as a framework for digital library design and evaluation. The model defines multiple components of the digital library, such as content, functionality, architecture, and users (Candela et al., 2007). Fuhr et al. (2001) developed a description scheme for digital library evaluation that defined four major components in the digital library domain: data/collection, system/technology, users, and usage. They explained the non-orthogonal relationships between those components and identified specific evaluation criteria for each component. Based on the DELOS Reference Model, Fuhr et al. (2007) conducted a large-scale survey of current digital library evaluation activities. The evaluation framework they suggested comprises a specific methodology for the classification of evaluation procedures.

Saracevic (2004) proposed a comprehensive digital library model for the purpose of evaluation. His model encompasses various aspects of the digital library in the evaluation, such as content, technology, interface, process and service, user, and context. Xie's (2006) framework has focused on users in the evaluation of digital libraries. It was designed to assess a digital library comprehensively from the perspectives of users. It identifies multiple dimensions of the criteria for digital library evaluation, consisting of usability, collection quality, service quality, system performance efficiency, and user-feedback solicitation. Lai et al. (2014) adopted the fuzzy analytic hierarchical process for assessing the interface of a digital library. Their evaluation framework determined the most important evaluation criteria as ease of use, searching, language, presentation, and design. Albertson (2015) developed a user-centered framework for visual digital libraries, which explored the three key components – the user, system, and interaction, as well as the surrounding domain and topics of a visual digital library, and further identified evaluation measures and methods for the relationships among these key components.

Noh (2010) proposed an array of evaluation criteria for digital libraries that covers various types of digital resources, such as databases, e-books, and e-journals. Her research came up with 11 evaluation criteria in 3 sectors: e-resource acquisition, e-resource use, and the e-resource environment. Noh (2011) used the data employment analysis method to measure the efficiency of digital resource use in academic libraries. Her evaluation model was

designed to assess how libraries' budget and input would be associated with the actual use of digital resources and research outputs. Gonçalves et al. (2004) proposed a unique model of digital libraries, the so-called "5S" model, which stands for streams, structures, spaces, scenarios, and societies. The 5S model can be used for the building and interpretation of a digital library taxonomy and analysis of digital library cases. Gonçalves et al. (2008) further developed a formal ontology for digital libraries based on the 5S model to create the fundamental concepts needed for digital library design and construction. Gonçalves et al. (2007) discussed the meaning of quality in digital libraries and proposed a quality model grounded on the 5S model with a range of criteria, including accessibility, accuracy, and completeness, among others.

Chowdhury (2014) developed a conceptual model for the sustainability of digital libraries. The model identifies three key components surrounding a digital library—data/content, users, and information and communications technology—as well as the major factors in relation to the economic, social, and environmental sustainability of digital libraries. His model suggested specific criteria for examining sustainability, such as funding models, impact assessment, user- and context-specific design, and sustainable preservation models. Kelly (2014) conducted a review of literature related to digital library evaluation published between 2004 and 2014. She identified a variety of key themes in digital library evaluation, ranging from usability testing and web statistics for data collection to altmetrics, the reuse of digital library materials, and cost–benefit analysis. Stiller and Petras (2018) reviewed evaluation studies of the Europeana Collections and explored various evaluation criteria in digital libraries based on a meta-analysis. They found that the evaluation studies adopted different types of methods, such as criteria-based evaluation, log file analysis, usability tests, and impact assessment. The criteria applied in the evaluation of the Europeana Collections consisted of data quality, usage statistics, accessibility, and coverage, among others.

Assessing the impact of digital libraries and evaluating the use and reuse of digital objects for educational and personal use are growing areas of research interest. The case studies evaluating the reuse of digital objects involve analysis of user requests, tracking the use of images on the Web using reverse image lookup (Reilly and Thompson, 2014, 2017), or examining the reuse of digital library objects in Wikipedia (Kelly, 2018). The "Measuring Reuse of Digital Objects" project, undertaken by members of the Digital Library Federation's Assessment Interest Group, focused on exploring the ways digital library materials are employed and reused (Kelly et al., 2018; O'Gara et al., 2018). The researchers made a distinction between use and reuse, defining use as interacting with digital objects within a digital repository and reuse as the ways digital library objects are repurposed in a known context (Kelly et al., 2018). Furthermore, the Assessment Interest Group research team explored the barriers and solutions in assessing the reuse of digital objects outside of digital library systems (O'Gara et al., 2018). Matusiak et al. (2019) expanded the notion of use and reuse and found instances of creative and transformative reuse in the academic context. These studies identify categories of reuse that can be applied in further evaluation studies of digital libraries.

Digital library evaluation measures

Different types of measures for various evaluation criteria have been devised and used for digital library evaluation. Researchers have developed evaluation tools with specific criteria/measures based on the DELOS Reference Model. Fuhr et al. (2001) proposed a digital library evaluation scheme involving four components (collection, technology, users, and usage) and associated evaluation criteria. They suggested measures including binary values, a restricted set of values, numeric values, and distribution. Tsakonas et al. (2004) advanced Fuhr et al.'s (2001) scheme. They identified evaluation criteria/sub-criteria for the dimensions of usability, usefulness, and performance, and then devised metrics to measure each sub-criterion. For example, they came up with measures for the criterion of user preference, including the number of accomplished tasks, number of abandoned tasks, and number of serendipitous accomplishments of tasks. Zhang (2010) adopted Saracevic's (2004) digital library evaluation framework for an empirical study. She measured different levels of the six criteria of Saracevic's framework using a Likert scale ranging from "Not significant at all" to "Extremely significant." Yan et al. (2014) examined different aspects of quality in digital libraries, such as information quality, system quality, and service quality. They created specific measurement items to compute different constructs of digital library quality. For example, the information quality of a digital library is measured by whether the digital library's items are up to date.

Researchers have generated measures to assess the usage of digital libraries. Noh (2010) suggested an array of specific quantitative indices for digital library usage, such as the annual number of the kinds of e-resources purchased per service recipient, the annual e-resource development cost per service recipient, and the number of e-book loans per service recipient, among others. Noh (2012) further elaborated the evaluation framework based on an analysis of the input and output of library resources. She measured the efficiency by calculating the ratio between the input and the output in a digital library context. Yang and Dawson (2018) adopted altmetrics as a means of measuring the impact of digital libraries. Examples of altmetrics data include downloads, reads, mentions, likes, views, and citations.

Digital library usability is another important area in digital library evaluation studies. Jeng (2005) developed a usability evaluation model for digital libraries to assess the effectiveness, efficiency, satisfaction, and learnability of a digital library. In her evaluation method, efficiency was computed by the time taken to complete a task and keystrokes and clicks, while effectiveness was measured by the proportion of correct answers. Learnability was inferred from the correctness of the task completion and time spent. Satisfaction was assessed from the perspectives of users' reactions using a Likert scale. Inal (2018) employed a heuristics approach for digital library usability evaluation, and worked with 57 subjects who had sufficient knowledge of the evaluation of a website with heuristics. They inspected the usability of a digital library with a set of predefined heuristics with different levels of severity, ranging from cosmetic to minor, major, and catastrophic. Some examples of heuristics items include visibility of the system's status, the match between the system and the real world, user control and freedom, and error prevention. Similarly, Ferati and Beyene (2017) adopted a heuristics inspection method for the evaluation of digital library accessibility. The set of heuristics items they developed consisted of 16 items, such as providing text alternatives for all non-text content, making it easier to see and hear content, and making all functionality available from a keyboard. Mune and Agee (2016) examined the accessibility of digital libraries for users with physical or learning disabilities. They identified specific measures for format and functionality. The measures included, for example, the inspection of a table of contents, a full-text search, paging forwards and backwards, and specifying a page number for the criterion of search/navigation.

Recently, researchers have paid attention to interactions between users and the interface in evaluating digital libraries. Albertson (2015) suggested a range of measures to evaluate visual digital libraries such as image and video digital libraries. As part of his digital library evaluation framework, he specifically identified measures to assess user interactions, such as complexity measured from a post-search survey, time observed from a system log, steps from experimental observation, and topic completion from experimental observation. Li and Liu (2019) examined users' interactions with digital libraries involving the factors of information resources, the interface, and tasks. In their experimental study, they assessed digital library performance in different ways, such as a reasonable page layout, the salience of topics, website organization, learnability, accessibility, and usefulness, among others. Also, they suggested several user side measures, such as task completion, satisfaction with the digital library design, success of the search, and satisfaction with the search process.

Other researchers have attempted to evaluate the software, metadata, and security of digital libraries. Gkoumas and Lazarinis (2015) inspected the current open-source software systems used for digital libraries. They adopted mostly binary measures to check if a certain system supported specific functions. For example, they examined whether certain software systems supported browsing, a simple search, an advanced search, and a full-text search. Park (2009) investigated criteria and measures for metadata quality evaluation, such as the degree of completeness, the correctness of the content of the data elements, and different levels of consistency. Huang et al. (2019) proposed an objective methodology for identifying and computing the factors relevant to the assessment of information security risks for digital libraries. They measured the grades of possibility of threats using a five-point scale ranging from very low (threat) to low, medium, high, and very high.

As outlined in this literature review section, researchers have made efforts to evaluate different aspects of digital libraries. They have identified specific criteria and corresponding measures to bring the reportable outcomes of evaluation. However, as far as we are aware, little research has been undertaken to comprehensively investigate the appropriateness of measures from the perspectives of stakeholders. Moreover, prior studies have suggested a variety of evaluation measures, but most have only attempted to apply those measures to a certain dimension of digital libraries.

Research questions and hypothesis

Prior studies have identified various criteria for digital library evaluation (e.g. Noh, 2010; Saracevic, 2004; Xie, 2006) and proposed different types of measures for some of the criteria in digital library evaluation practices (e.g. Jeng, 2005; Yan et al., 2014). Zhang (2010) investigated the importance of evaluation criteria and examined consistency among multiple stakeholders. However, existing research on digital library measures has only focused on one or a limited number of dimensions or criteria, and has not provided a comprehensive list of evaluation measures corresponding to each of the evaluation criteria. More importantly, there has been little effort to validate evaluation measures from the perspectives of scholars and practitioners. In order to fill these gaps in digital library evaluation, this study proposes the following two research questions: (1) What are the levels of appropriateness of measures corresponding to digital library evaluation criteria rated by digital library scholars and digital librarians? (2) What are the similarities and differences in perceptions of the appropriateness of measures corresponding to digital library evaluation criteria between digital library scholars and digital librarians? The associated hypothesis is: There is no significant difference in the distributions of the ratings of the appropriateness for each measure associated with a digital library evaluation criterion between digital library scholars and digital librarians.

Table 1. Demographic information.

	Category	Scholars	Digital librarian
Gender	Male	14 (46.7%)	13 (41.9%)
	Female	15 (50%)	18 (58.1%)
	No response	1 (3.3%)	
Age	31–40	6 (20%)	3 (9.7%)
	41–50	12 (40%)	10 (32.3%)
	51–60	7 (23.3%)	12 (38.7%)
	61+	5 (16.7%)	6 (19.4%)
Scholars' research areas	Digital libraries, information retrieval, metadata, digital humanities, human–computer interaction, digital preservation, information visualization		
Digital librarians' title	Digital Librarian, Digital Initiatives Librarian, Digital Projects Coordinator, Digital Collections Librarian, Digital Projects Manager, Head of Digital Collections, Head of Special Collections and Digital Initiatives		

Methodology

A thorough document analysis was conducted as the first step of this study to suggest an initial pool of dimensions and criteria with specific measures and data collection methods. Ten essential dimensions of digital libraries were identified: the collection, information organization, interface design, system and technology, effects on users, services, preservation, sustainability, user engagement, and context. The details of how the researchers came up with these dimensions from the document analysis can be found in a previously published article (Joo and Xie, 2013). Following this, two rounds of in-depth surveys were conducted to identify the importance of the evaluation criteria and the appropriateness of the measures from different stakeholders in digital libraries. The findings of the first round of the study on evaluation criteria were reported in the previous article (Xie et al., 2018). The present article concentrates on the results of the second round of the in-depth survey for the proposed research questions and associated hypothesis.

Sampling

The second-round survey was sent to the 61 respondents who participated in the first-round survey. All 30 of the digital library scholars and 31 digital librarians who were sent the survey participated in the second round of the study. The scholars group consisted of researchers who conduct digital library research with high citations and professors who have taught digital library courses. The former were identified based on the search results of the Web of Knowledge and Google Scholar, while the latter were selected from the websites of library and information science schools. In addition, 10 of the 30 scholars were international. The digital librarians were randomly selected from the top-200 US colleges which have operating digital libraries according to US News Rank, as well as from the partner libraries of this project.[1]

Table 1 presents the demographic information of the 30 digital library scholar and 31 digital librarian participants. For the scholars, it shows well-balanced proportions by professor rank. Their research areas cover different aspects of digital libraries, such as information retrieval, metadata, human–computer interaction, and preservation. On average, the digital librarians had approximately 8.48 (median = 8) years of experiences in digital-library-related services. Their official titles were diverse, including Digital Librarian, Digital Initiative Librarian, and Digital Collection Librarian. After checking the surveys, 29 scholar surveys and 27 digital librarian surveys were deemed valid for the appropriateness of measures ratings.

Data collection

Two rounds of in-depth surveys were administered. Since the survey in each round was quite comprehensive, it took each subject about 40–60 minutes to complete. In the first round, the importance of evaluation criteria was investigated based on a seven-point Likert scale. In this way, importance served as a key variable to rank the evaluation criteria. The purpose of the first round was to determine the importance of the digital library evaluation criteria under the 10 dimensions from the perspectives of different stakeholders and, moreover, to compare their similarities and differences (Xie et al., 2018).

The results from the first-round survey on the importance of the criteria were incorporated into the second round of the study, reported in this article. As all of the criteria were rated above 4.25 on a seven-point Likert scale by the scholars, librarians, and users in the first-round survey, none of the evaluation criteria were excluded from the second-round survey. The second-round survey investigated the appropriateness of measures to their corresponding criterion studied in the first round of the survey. For the second round, 174 measures were identified, since most of the criteria had multiple measures. Using a seven-point Likert scale, this study examined the appropriateness of measures to assess each evaluation criterion. The participants were offered definitions for each measure generated

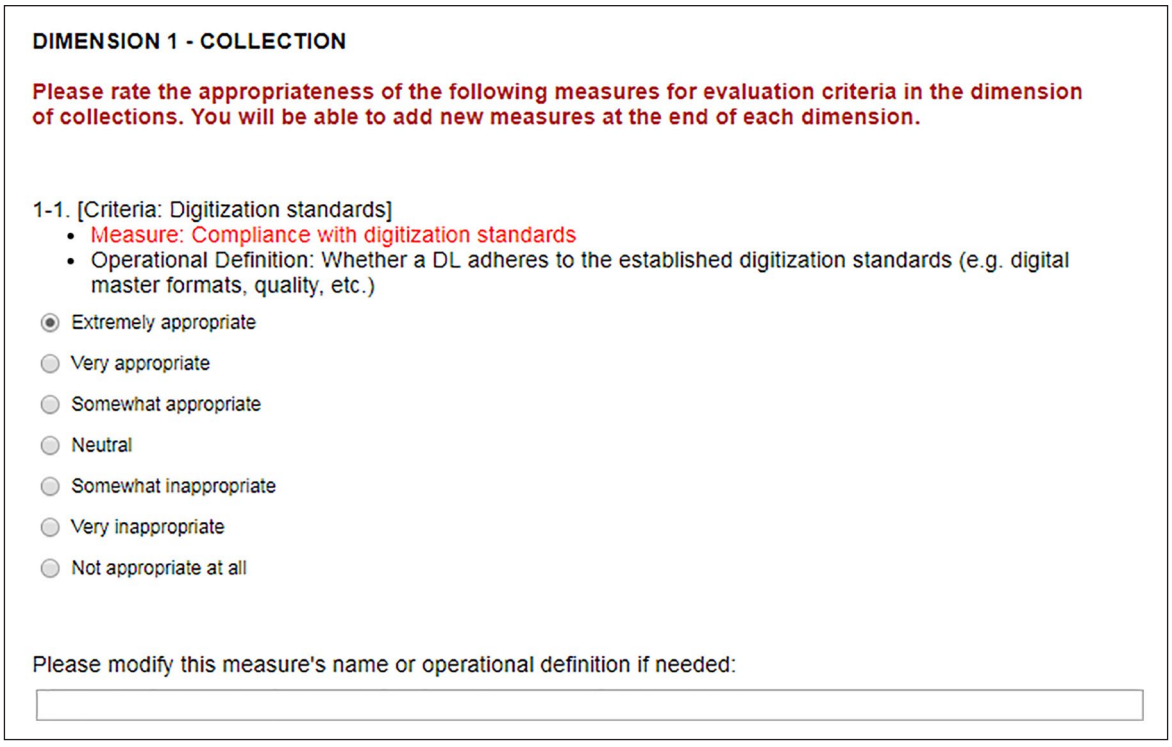
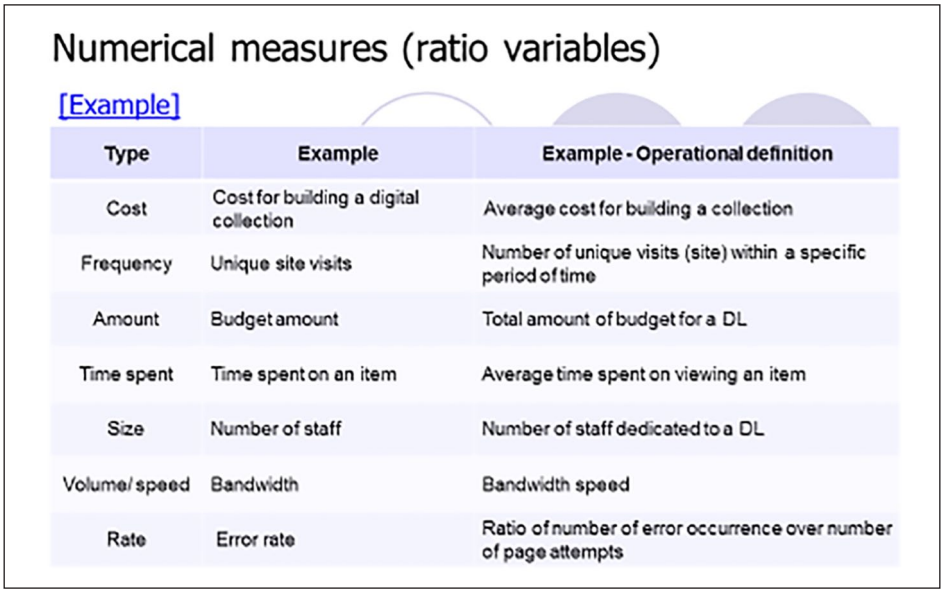


Figure 1. An example question.

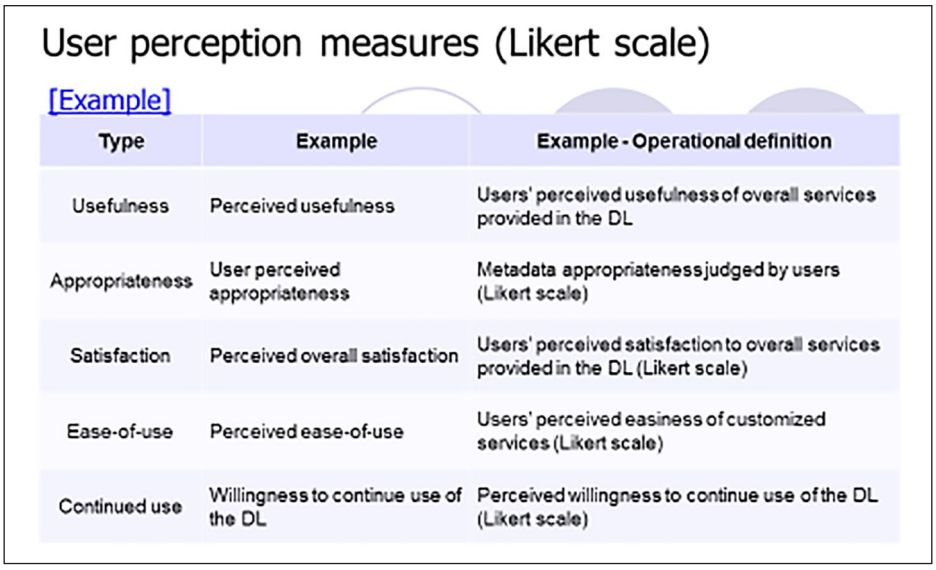


Type	Example	Example - Operational definition
Cost	Cost for building a digital collection	Average cost for building a collection
Frequency	Unique site visits	Number of unique visits (site) within a specific period of time
Amount	Budget amount	Total amount of budget for a DL
Time spent	Time spent on an item	Average time spent on viewing an item
Size	Number of staff	Number of staff dedicated to a DL
Volume/ speed	Bandwidth	Bandwidth speed
Rate	Error rate	Ratio of number of error occurrence over number of page attempts

Figure 2. Numerical measures.

by the document analysis and the authors' own recommendations. Figure 1 presents an example question about the appropriateness of digital library measurement for a digital library evaluation criterion in the second-round survey.

The authors came up with measures for each digital library evaluation criterion based on the relevant literature. Four types of measures representing both quantitative and qualitative approaches were employed as operational definitions: numerical measures (ratio variables; Figure 2), user perception measures (Likert scale; Figure 3), identification of types (Figure 4), and qualitative measures and dichotomous measures (Figure 5).

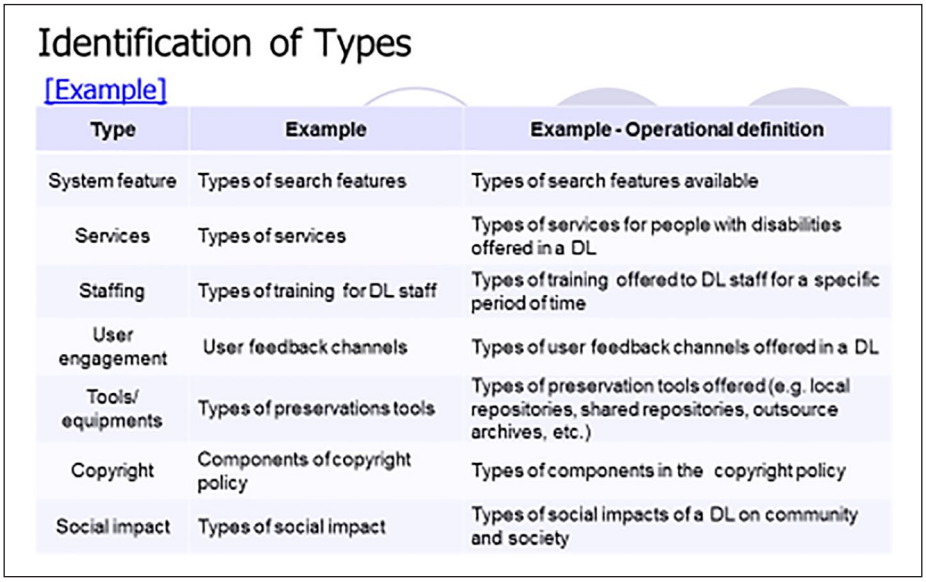


User perception measures (Likert scale)

[Example]

Type	Example	Example - Operational definition
Usefulness	Perceived usefulness	Users' perceived usefulness of overall services provided in the DL
Appropriateness	User perceived appropriateness	Metadata appropriateness judged by users (Likert scale)
Satisfaction	Perceived overall satisfaction	Users' perceived satisfaction to overall services provided in the DL (Likert scale)
Ease-of-use	Perceived ease-of-use	Users' perceived easiness of customized services (Likert scale)
Continued use	Willingness to continue use of the DL	Perceived willingness to continue use of the DL (Likert scale)

Figure 3. User perception measures.



Identification of Types

[Example]

Type	Example	Example - Operational definition
System feature	Types of search features	Types of search features available
Services	Types of services	Types of services for people with disabilities offered in a DL
Staffing	Types of training for DL staff	Types of training offered to DL staff for a specific period of time
User engagement	User feedback channels	Types of user feedback channels offered in a DL
Tools/ equipments	Types of preservations tools	Types of preservation tools offered (e.g. local repositories, shared repositories, outsource archives, etc.)
Copyright	Components of copyright policy	Types of components in the copyright policy
Social impact	Types of social impact	Types of social impacts of a DL on community and society

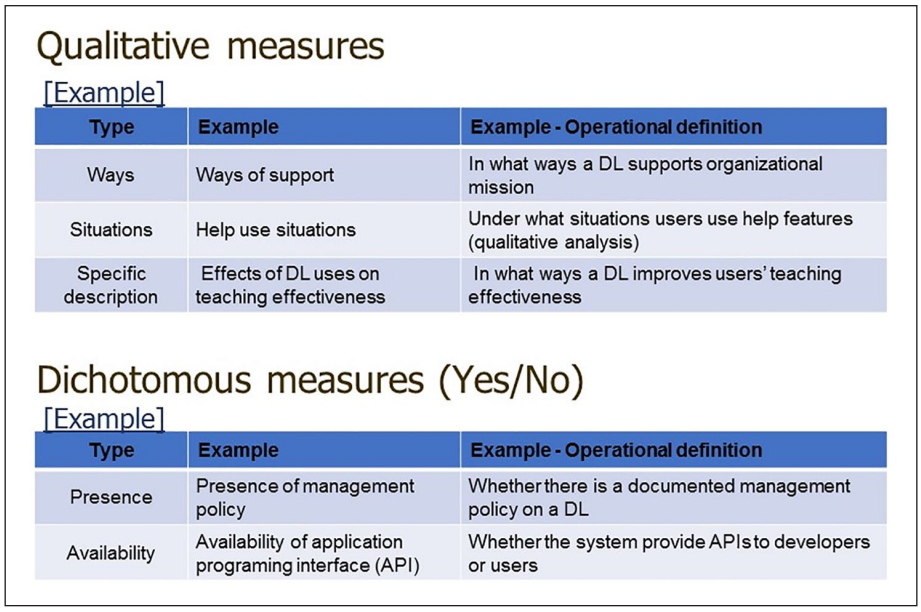
Figure 4. Identification of types.

Data analysis

As the majority of the data collected through the second-round survey contained numerical ratings, quantitative analysis was employed. Descriptive analysis was applied, including means and standard deviations (SDs), to identify the appropriateness of the digital library measures for the digital library evaluation criteria. Based on the ratings, the levels of appropriateness of the measures for the digital library evaluation criteria within each dimension were identified to validate the appropriateness of the measures to the corresponding criteria and to select the somewhat appropriate measures and inappropriate measures that needed to be further investigated. Moreover, the ratings of the evaluation measures between the scholars and librarians were compared in order to better understand the different perspectives of these two types of stakeholders. There was a certain portion of the measurement items that was not normally distributed. Thus, nonparametric Mann–Whitney tests were performed so that the similarities and differences between these two groups could be examined and tested.

Results

This results section presents the ratings of the appropriateness of the measures for each digital library evaluation



Qualitative measures

[Example]

Type	Example	Example - Operational definition
Ways	Ways of support	In what ways a DL supports organizational mission
Situations	Help use situations	Under what situations users use help features (qualitative analysis)
Specific description	Effects of DL uses on teaching effectiveness	In what ways a DL improves users' teaching effectiveness

Dichotomous measures (Yes/No)

[Example]

Type	Example	Example - Operational definition
Presence	Presence of management policy	Whether there is a documented management policy on a DL
Availability	Availability of application programing interface (API)	Whether the system provide APIs to developers or users

Figure 5. Qualitative measures and dichotomous measures.

criterion in 10 dimensions, as well as the Mann–Whitney test results of the comparison of the patterns of the ratings between the scholar and librarian groups. In each table, the criteria are organized based on their importance rated by different stakeholders in the first-round survey. In the tables below, the statistically significant results are in bold type for easy recognition. Detailed information about the significant results is reported in Appendix 1.

Dimension 1. Collections

Eighteen measures were suggested for thirteen criteria in this dimension (Table 2). The Mann–Whitney test results fail to reject the null hypothesis that there is no significant difference in the distributions in the ratings of the appropriateness of each measure associated with a digital library evaluation criterion in the dimension of "collections" between the scholar group and the librarian group. The participants rated four measures as "very appropriate" (6) or over: "compliance with digitization standards," "quality specification," "presence of resource reference information," and "data type." However, "potential user demographic data," "presence of diverse perspectives," and "currency of collections" were rated less than "somewhat appropriate" (5) to assess the corresponding criteria.

Dimension 2. Information organization

In the dimension of "information organization," 17 measures were suggested for 9 criteria (Table 3). Again, the Mann–Whitney test results fail to reject the null hypothesis that there is no significant difference in the distributions in the ratings of the appropriateness of each measure associated with a digital library evaluation criterion in the dimension of "information organization" between the scholar group and the librarian group. All of the measures were rated over 5 in terms of their appropriateness. There were two measures that were rated over 6: "compliance to the metadata standards" and "compliance to interoperability standards." The three least appropriate measures were "subject analysis," "depth of description," and "metadata elements used," with the former two measures corresponding to the criterion of "depth of metadata."

Dimension 3. Interface design

In the dimension of "interface design," 30 measures were identified for 9 criteria (Table 4). The Mann–Whitney test results fail to reject the null hypothesis that there is no significant difference in the ratings of the appropriateness of each measure associated with a digital library evaluation criterion in this dimension, except for "design consistency," between the scholar group and the librarian group. The median responses in the scholar group and librarian group for the measure of "design consistency" are 7 and 6, respectively. The distributions in the two groups differ significantly ($U = 230.5$, $p < .05$). Five measures were rated over 6 for their appropriateness: "search function usefulness," "types of search features," "search function ease of use," "overall ease of use," and "design consistency." However, two measures were perceived as less than "somewhat appropriate": "use of personalized features" and "use of help features."

Dimension 4. System/technology

In the fourth dimension, 15 measures for 10 criteria were identified (Table 5). The Mann–Whitney test results fail

Table 2. Appropriateness of evaluation measures: collections.

Criterion	Measure	Operational definition	Mean (SD)
Digitization standards	Compliance with digitization standards	Whether a digital library adheres to the established digitization standards	Scholar: 6.38 (0.775) Librarian: 6.26 (0.813) Overall: 6.32 (0.789)
Authority	Presence of resource reference information	Whether resource reference information for each item is available	Scholar: 6.11 (0.956) Librarian: 5.89 (0.751) Overall: 6.00 (0.861)
Cost	Cost of conversion	Average cost for converting a physical item to a digitized item	Scholar: 4.93 (1.307) Librarian: 5.56 (1.013) Overall: 5.23 (1.206)
	Cost of metadata	Average cost for creating metadata per record	Scholar: 5.25 (1.351) Librarian: 5.56 (0.934) Overall: 5.40 (1.164)
	Cost for building a digital collection	Average cost for building a collection	Scholar: 5.31 (1.466) Librarian: 5.52 (0.802) Overall: 5.41 (1.187)
Item quality	Quality specification	Technical specification for creating digitized objects	Scholar: 6.14 (0.932) Librarian: 6.11 (0.801) Overall: 6.13 (0.862)
Format compatibility	Types of access files	Types of access files used in the collection	Scholar: 5.97 (1.117) Librarian: 5.78 (0.934) Overall: 5.88 (1.028)
	Data type	Types of data used in the collection	Scholar: 5.36 (1.496) Librarian: 6.00 (1.000) Overall: 5.67 (1.306)
Audience	User demographic data	Whether types of user information data are collected	Scholar: 5.28 (1.334) Librarian: 4.81 (1.360) Overall: 5.05 (1.354)
	Potential user demographic data	Types of potential users and their demographic characteristics	Scholar: 5.00 (1.558) Librarian: 4.74 (1.023) Overall: 4.88 (1.322)
Scope/ coverage	Subject coverage	Number of topics in the digital library	Scholar: 5.50 (1.427) Librarian: 5.59 (1.185) Overall: 5.55 (1.303)
	Time span of coverage	Time period covered in the collections	Scholar: 5.66 (1.173) Librarian: 5.69 (0.884) Overall: 5.67 (1.037)
Contextual information	Presence of contextual information for collection	Whether there are secondary resources for digital collections to provide contextual information	Scholar: 5.50 (1.106) Librarian: 5.70 (0.775) Overall: 5.60 (0.955)
Completeness	Item size on specific topic	Number of items per topic	Scholar: 5.45 (1.429) Librarian: 5.04 (1.018) Overall: 5.25 (1.254)
Diversity	Presence of diverse perspectives	Whether a digital library contains diverse perspectives on a topic	Scholar: 4.93 (1.120) Librarian: 4.85 (1.231) Overall: 4.89 (1.165)
Size	Collection size	Number of digitized objects	Scholar: 5.69 (1.072) Librarian: 5.33 (1.441) Overall: 5.52 (1.265)
Collection development policy	Presence of collection development policy	Whether a digital library has a documented policy about collection development	Scholar: 5.89 (0.875) Librarian: 5.63 (1.043) Overall: 5.76 (0.962)
Currency	Currency of collections	Proportions of newly added collections	Scholar: 4.93 (1.412) Librarian: 5.00 (1.038) Overall: 4.96 (1.232)

to reject the null hypothesis that there is no significant difference in the distributions in the ratings of the appropriateness for each measure in this dimension, except "recall," between the two groups. The median responses in the scholar group and the librarian group for the measure of "recall" are 6 and 7, respectively. The distributions

Table 3. Appropriateness of evaluation measures: information organization.

Criterion	Measure	Operational definition	Mean (SD)
Appropriateness	User perceived appropriateness	Metadata appropriateness judged by users	Scholar: 5.66 (1.078) Librarian: 5.41 (1.118) Overall: 5.54 (1.095)
	Domain appropriateness	Metadata appropriateness judged by domain expert	Scholar: 5.86 (0.891) Librarian: 5.52 (1.051) Overall: 5.69 (0.979)
Accessibility to metadata	Ease of access to metadata	Users' perceived accessibility to metadata	Scholar: 5.82 (1.020) Librarian: 5.70 (0.869) Overall: 5.76 (0.942)
Metadata accuracy	Inaccurate data entry	Percentage of inaccurate data entry	Scholar: 6.03 (0.944) Librarian: 5.70 (0.993) Overall: 5.88 (0.974)
	Incorrect data value	Percentage of incorrect data value	Scholar: 6.00 (1.330) Librarian: 5.96 (0.871) Overall: 5.98 (1.118)
Metadata standards	Type of metadata standards	Whether a digital library adheres to the selected metadata standard	Scholar: 5.93 (1.215) Librarian: 5.78 (1.013) Overall: 5.85 (1.113)
	Compliance to the metadata standards	Proportion of accurately mapped elements compared with the selected metadata scheme	Scholar: 6.26 (0.712) Librarian: 5.93 (0.917) Overall: 6.09 (0.830)
Consistency	Metadata schema consistency	To what extent the selected metadata scheme is used for data input across collections in a digital library	Scholar: 5.78 (1.050) Librarian: 5.85 (0.784) Overall: 5.81 (0.921)
	Metadata element consistency	To what extent the selected metadata elements are used for data input across collections in a digital library	Scholar: 5.96 (0.793) Librarian: 5.93 (0.730) Overall: 5.95 (0.756)
Comprehensiveness	Completed metadata	Average number of metadata fields filled per record	Scholar: 5.43 (1.069) Librarian: 4.85 (1.262) Overall: 5.15 (1.193)
	Metadata elements used	Proportions of metadata elements used compared with the selected metadata scheme	Scholar: 5.11 (1.257) Librarian: 5.07 (1.107) Overall: 5.09 (1.175)
Depth of metadata	Subject analysis	Average number of subject terms per record	Scholar: 5.07 (1.215) Librarian: 4.93 (0.958) Overall: 5.00 (1.089)
	Depth of description	Length of item description	Scholar: 5.21 (1.166) Librarian: 4.78 (1.050) Overall: 5.00 (1.122)
Metadata interoperability	Compliance to interoperability standards	Whether a digital library complies with interoperability standards	Scholar: 6.07 (0.766) Librarian: 6.07 (0.917) Overall: 6.07 (0.836)
Controlled vocabulary	Presence of controlled vocabularies	Whether a digital library uses controlled vocabularies in organizing objects	Scholar: 5.86 (0.765) Librarian: 6.04 (0.649) Overall: 5.95 (0.705)
	Ease of access to controlled vocabularies	Users' perceived accessibility	Scholar: 5.54 (1.105) Librarian: 5.62 (0.804) Overall: 5.57 (0.964)
	Presence of controlled vocabularies	Whether a digital library offers controlled vocabularies	Scholar: 5.36 (1.224) Librarian: 5.63 (0.629) Overall: 5.49 (0.979)

in the two groups differ significantly ($U = 202, p < .05$). While all of the measures in this dimension were rated above 5, five measures were perceived to be above "very appropriate": "precision," "response time to search results," "system failure," "recall," and "system response time."

Table 4. Appropriateness of evaluation measures: interface design.

Criterion	Measure	Operational definition	Mean (SD)
Search function	Types of search features	Types of search features available	Scholar: 6.28 (0.797) Librarian: 6.27 (0.724) Overall: 6.27 (0.757)
	Usefulness	Users' perceived usefulness	Scholar: 6.45 (0.827) Librarian: 6.12 (0.864) Overall: 6.29 (0.854)
	Ease of use	Users' perceived ease of use	Scholar: 6.30 (0.953) Librarian: 6.15 (0.864) Overall: 6.22 (0.904)
	Use of search features	Average frequency of and time spent on search feature uses in a session	Scholar: 5.28 (1.099) Librarian: 5.59 (1.010) Overall: 5.43 (1.059)
Browsing function	Browsing access points and paths	Types of browsing access points and paths available	Scholar: 5.79 (1.048) Librarian: 6.00 (0.734) Overall: 5.89 (0.908)
	Organization of browsing structure	Expert assessment of logic and quality of browsing structure	Scholar: 5.89 (0.956) Librarian: 5.78 (0.751) Overall: 5.84 (0.856)
	Usefulness	Users' perceived usefulness	Scholar: 6.11 (0.956) Librarian: 5.85 (0.818) Overall: 5.98 (0.892)
	Ease of use	Users' perceived ease of use	Scholar: 5.93 (1.016) Librarian: 6.00 (0.832) Overall: 5.96 (0.922)
	Use of browsing features	Average frequency of and time spent on browsing feature uses in a session	Scholar: 5.14 (1.380) Librarian: 5.41 (0.844) Overall: 5.27 (1.146)
Navigation	Navigation features	Types of navigation features available	Scholar: 5.62 (1.237) Librarian: 6.00 (0.784) Overall: 5.80 (1.052)
	Usefulness	Users' perceived usefulness	Scholar: 5.93 (1.328) Librarian: 5.96 (0.759) Overall: 5.94 (1.071)
	Ease of use	Users' perceived ease of use	Scholar: 5.79 (1.292) Librarian: 6.04 (0.854) Overall: 5.91 (1.100)
	Use of navigation features	Average frequency of and time spent on navigation feature uses in a session	Scholar: 5.07 (1.386) Librarian: 5.59 (0.797) Overall: 5.33 (1.156)
Intuitive operation	Overall ease of use	Users' perceived ease of use to operate the interface of a digital library	Scholar: 6.14 (1.079) Librarian: 6.12 (0.909) Overall: 6.13 (0.991)
Consistency	**Design consistency**	**Consistency in fonts, layout, menus, colors, etc.**	**Scholar: 6.36 (0.780)** **Librarian: 5.81 (0.801)** **Overall: 6.09 (0.830)**
	Consistency from users' perspective	Users' perceived consistency	Scholar: 5.82 (1.278) Librarian: 5.67 (0.877) Overall: 5.75 (1.092)
Help function	Types of help features	Types of help features available	Scholar: 5.52 (1.153) Librarian: 5.78 (0.934) Overall: 5.64 (1.052)
	Usefulness	Users' perceived usefulness	Scholar: 5.79 (0.978) Librarian: 5.81 (0.962) Overall: 5.80 (0.961)
	Ease of use	Users' perceived ease of use	Scholar: 5.78 (1.013) Librarian: 5.77 (0.992) Overall: 5.77 (0.993)

(Continued)

Table 4. (Continued)

Criterion	Measure	Operational definition	Mean (SD)
	Use of help features	Average frequency of and time spent on help feature uses in a session	Scholar: 4.79 (1.346) Librarian: 5.15 (0.989) Overall: 4.96 (1.19)
	Help use situations	Under what situations users use help features	Scholar: 5.63 (1.079) Librarian: 5.63 (0.926) Overall: 5.63 (0.996)
Visual appeal	Visual aesthetics	Users' perception of the interface aesthetics	Scholar: 5.75 (1.041) Librarian: 5.63 (0.839) Overall: 5.69 (0.940)
User control	Types of user control features	Types of user control features available	Scholar: 5.62 (1.147) Librarian: 5.52 (0.849) Overall: 5.57 (1.006)
	Usefulness	Users' perceived usefulness	Scholar: 5.93 (1.294) Librarian: 5.59 (0.844) Overall: 5.76 (1.088)
	Ease of use	Users' perceived ease of use	Scholar: 5.75 (1.378) Librarian: 5.73 (0.778) Overall: 5.74 (1.119)
	Use of user control feature	Average frequency of and time spent on using each type of user control feature in a session	Scholar: 4.97 (1.295) Librarian: 5.30 (0.869) Overall: 5.13 (1.113)
Personalization feature	Types of personalization features	Types of personalization features	Scholar: 5.34 (1.078) Librarian: 5.07 (1.072) Overall: 5.21 (1.074)
	Usefulness	Users' perceived usefulness	Scholar: 5.55 (1.242) Librarian: 5.30 (0.993) Overall: 5.43 (1.126)
	Ease of use	Users' perceived ease of use	Scholar: 5.50 (1.208) Librarian: 5.37 (0.967) Overall: 5.43 (1.083)
	Use of personalization features	Average frequency of and time spent on personalization feature uses in a session	Scholar: 4.78 (1.086) Librarian: 5.08 (0.891) Overall: 4.92 (0.997)

Table 5. Appropriateness of evaluation measures: system/technology.

Criterion	Measure	Operational definition	Mean (SD)
Retrieval effectiveness	**Recall**	**Recall = number of relevant items retrieved / number of relevant items in a digital library**	**Scholar: 5.50 (1.552)** **Librarian: 6.56 (0.577)** **Overall: 6.02 (1.284)**
	Precision	Precision = number of relevant items retrieved / number of retrieved items	Scholar: 6.00 (1.333) Librarian: 6.48 (0.580) Overall: 6.24 (1.053)
	Aspectual recall	Ratio of aspects of the search topic identified in the documents saved by the subject to the total number of aspects of the topic	Scholar: 5.17 (1.441) Librarian: 5.85 (0.967) Overall: 5.49 (1.275)
Retrieval efficiency	Response time to search results	Response time to present search results after a search request submitted	Scholar: 6.14 (1.044) Librarian: 6.19 (0.786) Overall: 6.16 (0.918)
Reliability	System failure	Number of system failures in a specific period of time	Scholar: 6.19 (0.879) Librarian: 6.11 (0.698) Overall: 6.15 (0.787)
Server performance	Bandwidth	Bandwidth speed	Scholar: 5.75 (1.110) Librarian: 5.96 (0.808) Overall: 5.85 (0.970)

(Continued)

Table 5. (Continued)

Criterion	Measure	Operational definition	Mean (SD)
	Traffic	The volume of total traffic accessing a digital library site	Scholar: 5.59 (1.118) Librarian: 5.93 (0.829) Overall: 5.76 (0.989)
Response time	System response time	End-to-end response time after a page request is made	Scholar: 5.96 (1.347) Librarian: 6.04 (0.808) Overall: 6.00 (1.106)
Fit to task	Perceived as fit to task	To what extent a user perceives the appropriateness of the system to carry out their search task	Scholar: 5.56 (1.188) Librarian: 5.96 (0.898) Overall: 5.76 (1.063)
Speed of page loading	Page loading speed	Average downloading speed per page	Scholar: 5.39 (1.548) Librarian: 5.96 (0.759) Overall: 5.67 (1.248)
Integrated search	Search across collections	Whether a digital library provides an integrated search function across multiple collections	Scholar: 5.75 (1.175) Librarian: 6.19 (0.879) Overall: 5.96 (1.053)
Error rate and correction	Error rate	Ratio of number of error occurrences over number of page attempts	Scholar: 5.57 (1.425) Librarian: 6.00 (0.764) Overall: 5.77 (1.171)
	Error correction rate	Ratio of corrected errors compared with errors encountered	Scholar: 5.59 (1.366) Librarian: 5.80 (0.957) Overall: 5.69 (1.181)
System connectivity	Compatible with other types of systems	Technical ability to connect to other types of systems	Scholar: 5.69 (1.258) Librarian: 5.89 (0.974) Overall: 5.79 (1.116)
	Easiness of connection	Expert assessment of easiness of connection to other systems	Scholar: 5.64 (0.989) Librarian: 5.59 (0.844) Overall: 5.62 (0.913)

Dimension 5. Effects on users

In the dimension of "effects on users," 11 measures were recommended for 5 evaluation criteria (Table 6). The Mann–Whitney test results fail to reject the null hypothesis in most measures except for "digital library uses for research" and "digital library uses for teaching." The median responses for the measure of "digital library uses for research" are 6 for both groups. The distributions in the two groups differ significantly ($U = 251.5$, $p < .05$). The median responses in the scholar group and the librarian group for the measure of "digital library uses for teaching" are 5 and 6, respectively. The distributions in the two groups differ significantly ($U = 235$, $p < .05$). In this dimension, no measure scored over 6. The four top-ranked measures in terms of appropriateness are "willingness to continue use of the digital library," "digital library uses for teaching," "digital library uses for research," and "effects of digital library uses on research." At the same time, two measures were rated at exactly 5: "attitude change after digital library uses" and "perceived information literacy/skill change."

Dimension 6. Services

In the dimension of "services," 25 measures were proposed for 9 criteria (Table 7). The Mann–Whitney test results fail to reject the null hypothesis that there is no significant difference in the distributions in the ratings of the appropriateness of each measure associated with a digital library evaluation criterion in this dimension. There were two measures that were rated over 6: "perceived usefulness" and "perceived satisfaction." These two measures are frequently used in service evaluation and they were also selected as highly appropriate measures in the context of digital libraries. However, the ratings for "staff accessible hours," "number of reference services provided," and "types of services–uniqueness," were lower than 5.

Dimension 7. Preservation

In the dimension of "preservation," nine measures were recommended for five criteria (Table 8). The Mann–Whitney test results fail to reject the null hypothesis that there is no significant difference in the distributions in the ratings of the appropriateness of each measure associated with a digital library evaluation criterion between the groups in this dimension. Among all the measures, "exporting capability," "presence of preservation policy," and "migratable data type" were rated above "very appropriate." The rest of the measures were regarded as "somewhat appropriate."

Table 6. Appropriateness of evaluation measures: effects on users.

Criterion	Measure	Operational definition	Mean (SD)
Research productivity	**Digital library uses for research**	**Frequency of digital library uses for research purposes**	**Scholar: 5.34 (0.936)** **Librarian: 5.96 (0.759)** **Overall: 5.64 (0.903)**
	Effects of digital library uses on research	In what ways a digital library enhances a user's research productivity	Scholar: 5.38 (1.347) Librarian: 5.93 (0.675) Overall: 5.64 (1.103)
	Research productivity change	To what extent a digital library enhances a user's research productivity	Scholar: 5.39 (1.449) Librarian: 5.33 (0.961) Overall: 5.36 (1.223)
Knowledge change	Perceived domain knowledge change	Perceived increase of domain knowledge after using the digital library	Scholar: 5.36 (1.311) Librarian: 5.33 (0.679) Overall: 5.35 (1.04)
	Domain knowledge change after digital library uses	Domain knowledge change between the pre-test and post-test	Scholar: 5.00 (1.074) Librarian: 5.15 (0.662) Overall: 5.07 (0.887)
Instructional effectiveness	**Digital library uses for teaching**	**Frequency of digital library uses in teaching**	**Scholar: 5.36 (0.780)** **Librarian: 5.96 (0.854)** **Overall: 5.65 (0.865)**
	Effects of digital library uses on teaching effectiveness	In what ways a digital library improves users' teaching effectiveness	Scholar: 5.00 (1.387) Librarian: 5.70 (0.823) Overall: 5.35 (1.184)
Perception of digital libraries	Attitude change after digital library uses	To what extent users change attitude toward digital libraries	Scholar: 4.90 (1.543) Librarian: 5.11 (0.892) Overall: 5.00 (1.265)
	Willingness to continue use of the digital library	Perceived willingness to continue use of a digital library	Scholar: 5.96 (0.808) Librarian: 5.85 (0.818) Overall: 5.91 (0.807)
Information literacy/skill change	Perceived information literacy/skill change	Perceived improvement of information literacy/skill after using a digital library	Scholar: 4.86 (1.382) Librarian: 5.15 (0.718) Overall: 5.00 (1.112)
	Change of information literacy/skill	Change between the pre-test and post-test	Scholar: 4.93 (0.940) Librarian: 5.15 (0.718) Overall: 5.04 (0.838)

Table 7. Appropriateness of evaluation measures: services.

Criterion	Measure	Operational definition	Mean (SD)
Usefulness	Perceived usefulness	Users' perceived usefulness of overall services provided in a digital library	Scholar: 6.28 (0.702) Librarian: 6.00 (0.877) Overall: 6.14 (0.796)
	Ways of usefulness	In what ways digital library services help users	Scholar: 5.57 (1.034) Librarian: 5.59 (0.888) Overall: 5.58 (0.956)
Satisfaction	Perceived satisfaction	Users' perceived satisfaction with overall services provided in the digital library	Scholar: 6.04 (1.232) Librarian: 6.07 (0.781) Overall: 6.05 (1.026)
Services for users with disabilities	Types of services	Types of services for people with disabilities	Scholar: 5.96 (0.793) Librarian: 5.93 (0.958) Overall: 5.95 (0.870)
	Usefulness	Disabled users' perceived usefulness	Scholar: 5.86 (0.848) Librarian: 5.96 (0.854) Overall: 5.91 (0.845)
	Ease of use	Disabled users' perceived ease of use for the services	Scholar: 5.93 (0.842) Librarian: 5.96 (0.854) Overall: 5.95 (0.840)

(Continued)

Table 7. (Continued)

Criterion	Measure	Operational definition	Mean (SD)
	Frequency of service uses by people with disabilities	Frequency of each type of services used by people with disabilities in a specific time period	Scholar: 4.74 (1.347) Librarian: 5.41 (1.047) Overall: 5.07 (1.242)
Reliability	Reliability of services	Users' perceived reliability	Scholar: 5.90 (1.263) Librarian: 5.96 (0.854) Overall: 5.93 (1.076)
Responsiveness	Perceived responsiveness	Service responsiveness rated by users	Scholar: 5.67 (1.301) Librarian: 5.89 (0.698) Overall: 5.78 (1.04)
Types of services	Types of user services	Types of user services offered in a digital library	Scholar: 5.45 (1.021) Librarian: 5.59 (0.844) Overall: 5.52 (0.934)
	Usefulness	Users' perceived usefulness for each type of service	Scholar: 5.89 (0.956) Librarian: 5.63 (0.839) Overall: 5.76 (0.902)
	Ease of use	Users' perceived ease of use for each type of service	Scholar: 5.85 (0.949) Librarian: 5.67 (0.877) Overall: 5.76 (0.91)
	Uniqueness	Number of unique services offered in a digital library	Scholar: 4.71 (1.630) Librarian: 5.07 (1.357) Overall: 4.89 (1.499)
	Frequency of service uses	Frequency of each type of service used in a specific time period	Scholar: 4.89 (1.281) Librarian: 5.15 (0.881) Overall: 5.02 (1.101)
Accessibility to managerial staff	Perceived availability of staff	Users' perceived staff availability	Scholar: 5.00 (1.363) Librarian: 5.48 (0.753) Overall: 5.23 (1.128)
	Staff accessible hours	Number of hours users can access digital library staff	Scholar: 4.71 (1.329) Librarian: 4.93 (0.917) Overall: 4.82 (1.140)
Reference services	Types of reference services	Types of reference services offered in a digital library	Scholar: 5.36 (0.989) Librarian: 5.22 (1.219) Overall: 5.29 (1.100)
	Usefulness	Users' perceived usefulness of reference services offered in a digital library	Scholar: 5.50 (1.000) Librarian: 5.33 (1.240) Overall: 5.42 (1.117)
	Ease of use	Users' perceived ease of use of using reference services	Scholar: 5.59 (1.010) Librarian: 5.37 (1.245) Overall: 5.48 (1.128)
	Number of reference services provided	Number of reference services provided in a specific period of time	Scholar: 4.79 (1.166) Librarian: 4.96 (1.400) Overall: 4.87 (1.277)
	Response time to digital library reference requests	Average response time to a reference request for digital library resources	Scholar: 5.50 (1.139) Librarian: 5.35 (1.198) Overall: 5.43 (1.159)
Customized services	Types of customized services	Types of customized services offered by a digital library	Scholar: 5.29 (1.013) Librarian: 5.48 (0.935) Overall: 5.38 (0.972)
	Usefulness	Users' perceived usefulness of customized services	Scholar: 5.50 (0.923) Librarian: 5.37 (0.839) Overall: 5.44 (0.877)
	Ease of use	Users' perceived easiness of customized services	Scholar: 5.43 (1.034) Librarian: 5.30 (0.869) Overall: 5.36 (0.95)
	Use of customized services	Frequency of each type of customized service used in a specific period of time	Scholar: 4.86 (1.113) Librarian: 5.15 (1.027) Overall: 5.00 (1.072)

Table 8. Appropriateness of evaluation measures: preservation.

Criterion	Measure	Operational definition	Mean (SD)
Ability to migrate	Migratable data type	Types of data that can be migrated to digital library	Scholar: 5.79 (1.449) Librarian: 6.33 (0.734) Overall: 6.05 (1.182)
	Exporting capability	Whether a digital library has a function to export data in different formats for preservation	Scholar: 6.21 (0.902) Librarian: 6.44 (0.847) Overall: 6.32 (0.876)
Preservation policy	Presence of preservation policy	Whether a digital library has a documented policy regarding preservation practices	Scholar: 6.03 (0.865) Librarian: 6.33 (0.832) Overall: 6.18 (0.855)
	Components of preservation policy	Types of components in the preservation policy	Scholar: 5.32 (1.389) Librarian: 5.85 (0.818) Overall: 5.58 (1.166)
	Strategies of preservation	Types of strategies presented in the preservation policy	Scholar: 5.38 (1.347) Librarian: 5.96 (0.854) Overall: 5.66 (1.164)
Preservation infrastructure	Types of preservations tools	Types of preservation tools offered	Scholar: 5.50 (1.427) Librarian: 6.00 (0.784) Overall: 5.75 (1.174)
Institutional support	Types of support	Types of support offered by the institution	Scholar: 5.79 (1.292) Librarian: 5.96 (1.091) Overall: 5.88 (1.192)
	Level of support	The extent of support offered by the institution	Scholar: 5.71 (1.049) Librarian: 6.07 (0.829) Overall: 5.89 (0.956)
Cost per record	Preservation cost per record	Average cost for preserving a record	Scholar: 5.26 (1.130) Librarian: 5.31 (1.320) Overall: 5.28 (1.215)

Dimension 8. Administration

In the "administration" dimension, 22 measures were identified for 10 evaluation criteria (Table 9). The Mann–Whitney test results fail to reject the null hypothesis that there is no significant difference in the ratings of the appropriateness of each measure associated with a digital library evaluation criterion in the dimension of "administration" between the two groups, except for "staff hours on digital libraries" and "student hours on digital libraries." The median responses in the scholar group and the librarian group for the measure "staff hours on digital libraries" are 5 and 6, respectively. The distributions in the two groups differ significantly ($U = 239.5$, $p < .05$). The median responses in the scholar group and the librarian group for the measure "student hours on digital libraries" are 4 and 5, respectively. The distributions in the two groups differ significantly ($U = 196$, $p < .05$). "Presence of copyright policy" and "presence of sustainability plans" turned out to be the two "very appropriate" measures for their associated criteria. Conversely, five measures were rated below "somewhat appropriate," including "proportion of DL budget," "student hours on digital libraries," "number/amount of grant/fundraising," "number/amount of grant/fundraising received," and "frequency of marketing/promotion activities."

Dimension 9. User engagement

In the dimension of "user engagement," 13 measures were recommended for 5 criteria (Table 10). The Mann–Whitney test results fail to reject the null hypothesis that there is no significant difference in the ratings of the appropriateness of each measure associated with a digital library evaluation criterion in this dimension between the groups, except for "unique page visits." The median responses for the measure of "unique page visits" are 6 for both groups. The distributions in the two groups differ significantly ($U = 245$, $p < .05$). In particular, five measures were suggested for the criterion of "site visits": "frequency of site visits," "session length," "frequency of page visits," "unique site visits," and "unique page visits." In this dimension, most of the measures were related to measuring resource usage in digital libraries. All of the measures were rated as "somewhat important."

Dimension 10. Context

Finally, 14 measures were suggested for 7 criteria in the dimension of "context" (Table 11). The Mann–Whitney test results fail to reject the null hypothesis that there is no significant difference in the ratings of the appropriateness

Table 9. Appropriateness of evaluation measures: administration.

Criterion	Measure	Operational definition	Mean (SD)
Copyright	Presence of copyright policy	Whether a digital library has a documented policy regarding copyright management	Scholar: 6.32 (0.772) Librarian: 6.22 (0.847) Overall: 6.27 (0.804)
	Components of copyright policy	Types of components in the copyright policy	Scholar: 5.39 (1.370) Librarian: 5.63 (1.006) Overall: 5.51 (1.200)
Budget	Budget amount	Total amount of budget for a digital library	Scholar: 5.26 (1.583) Librarian: 5.41 (1.421) Overall: 5.33 (1.492)
	Proportion of digital library budget	Proportion of digital library budget over total library budget	Scholar: 4.58 (1.748) Librarian: 5.37 (1.214) Overall: 4.98 (1.538)
	Distribution of digital library budget	Distribution of budget for different components of a digital library	Scholar: 4.81 (1.545) Librarian: 5.30 (1.103) Overall: 5.06 (1.352)
Planning	Presence of strategic plans	Whether there are documented strategic plans on a digital library	Scholar: 5.78 (0.974) Librarian: 5.67 (1.074) Overall: 5.72 (1.017)
	Components of plans	Components of strategic plans	Scholar: 5.22 (1.188) Librarian: 5.63 (0.926) Overall: 5.43 (1.075)
Staffing	Number of staff	Number of staff dedicated to a digital library	Scholar: 5.11 (1.188) Librarian: 5.63 (1.149) Overall: 5.37 (1.186)
	Staff hours on digital libraries	**Number of professional staff hours dedicated to a digital library**	**Scholar: 4.77 (1.306)** **Librarian: 5.52 (1.156)** **Overall: 5.15 (1.277)**
	Student hours on digital libraries	**Number of student worker hours dedicated to a digital library**	**Scholar: 4.25 (1.236)** **Librarian: 5.30 (1.031)** **Overall: 4.76 (1.247)**
Staff training	Types of training for digital library staff	Types of training offered to digital library staff for a specific period of time	Scholar: 4.96 (1.372) Librarian: 5.44 (0.847) Overall: 5.20 (1.155)
	Resources	Amount of resources for training allocated to a digital library	Scholar: 5.00 (1.177) Librarian: 5.30 (0.993) Overall: 5.15 (1.089)
Marketing/ promotion	Marketing/promotion methods	Types of marketing methods used for promoting a digital library	Scholar: 4.73 (1.614) Librarian: 5.41 (0.971) Overall: 5.08 (1.357)
	Frequency of marketing/ promotion activities	Frequency of each type of marketing activity taking place for a digital library in a specific period of time	Scholar: 4.67 (1.387) Librarian: 5.07 (0.958) Overall: 4.87 (1.198)
	Recognition of digital libraries	Number of people aware of a digital library	Scholar: 4.96 (1.483) Librarian: 5.33 (0.784) Overall: 5.15 (1.183)
Assessment	Frequency of assessment	Frequency of digital library assessment in a specific period of time	Scholar: 5.54 (1.067) Librarian: 5.41 (0.931) Overall: 5.47 (0.992)
	Dimensions of assessment	Dimensions of a digital library included in the assessment	Scholar: 5.65 (1.263) Librarian: 5.52 (1.189) Overall: 5.58 (1.216)
Management policy	Presence of management policy	Whether there is a documented management policy on a digital library	Scholar: 5.73 (0.874) Librarian: 5.41 (0.844) Overall: 5.57 (0.866)
	Components of management policy	Types of components in the management policy related to the digital library	Scholar: 5.23 (1.210) Librarian: 5.19 (1.111) Overall: 5.21 (1.15)

(Continued)

Table 9. (Continued)

Criterion	Measure	Operational definition	Mean (SD)
Grant/ fundraising	Number/amount of grant/fundraising	Total number/amount of grant/fundraising for digital libraries in a specific period of time	Scholar: 4.62 (1.472) Librarian: 5.07 (0.874) Overall: 4.85 (1.215)
	Number/amount of grant/fundraising received	Number/amount of grant/fundraising received in a specific period of time	Scholar: 4.73 (1.458) Librarian: 5.04 (0.940) Overall: 4.89 (1.219)
Sustainability plan	Presence of sustainability plans	Whether there are documented sustainability plans on a digital library	Scholar: 6.15 (0.770) Librarian: 5.93 (0.874) Overall: 6.04 (0.823)

Table 10. Appropriateness of evaluation measures: user engagement.

Criterion	Measure	Operational definition	Mean (SD)
Digital object use	Items viewed	Number of items viewed	Scholar: 5.78 (1.121) Librarian: 6.11 (1.050) Overall: 5.94 (1.089)
	Time spent on an item	Average time spent on viewing an item	Scholar: 5.44 (1.155) Librarian: 5.59 (1.338) Overall: 5.52 (1.240)
	Item downloading	Number of items downloaded	Scholar: 6.00 (1.095) Librarian: 5.78 (1.050) Overall: 5.89 (1.068)
User feedback	User feedback channels	Types of user feedback channels offered in a digital library	Scholar: 5.81 (0.736) Librarian: 5.85 (0.718) Overall: 5.83 (0.720)
	Quantity of user feedback	Number of user feedback submitted	Scholar: 5.30 (1.203) Librarian: 5.48 (1.014) Overall: 5.39 (1.106)
Site visit	Frequency of site visits	Number of site visits within a specific period of time	Scholar: 5.85 (1.027) Librarian: 5.93 (1.035) Overall: 5.89 (1.022)
	Session length	Average time spent on a digital library in a session	Scholar: 5.22 (1.502) Librarian: 5.78 (1.219) Overall: 5.50 (1.384)
	Frequency of page visits	Number of page visits within a specific period of time	Scholar: 5.52 (1.282) Librarian: 5.93 (0.997) Overall: 5.72 (1.156)
	Unique site visits	Number of unique visits (site) within a specific period of time	Scholar: 5.64 (1.283) Librarian: 6.07 (0.958) Overall: 5.85 (1.145)
	Unique page visits	**Number of unique visits (page) within a specific period of time**	**Scholar: 5.48 (1.252)** **Librarian: 6.15 (0.907)** **Overall: 5.81 (1.134)**
Integration with external applications	Compatibility with external applications	Types of external applications that are compatible with the digital library	Scholar: 5.50 (1.393) Librarian: 5.59 (0.931) Overall: 5.55 (1.170)
	Ease of integration	Degree of easiness to integrate a digital library to external application assessed by experts	Scholar: 5.44 (1.086) Librarian: 5.48 (0.802) Overall: 5.46 (0.946)
User/community knowledge contribution	Types of user/community knowledge contribution	Types of user/community knowledge contribution channels available in a digital library	Scholar: 5.52 (0.935) Librarian: 5.30 (0.953) Overall: 5.41 (0.942)

of each measure associated with a digital library evaluation criterion in the dimension of "context" between the two groups, except for "presence of ethics guidelines" and "level of social impact." The median responses in the scholar group and the librarian group for the measure of "presence of ethics guidelines" are 6 and 5, respectively.

Table 11. Appropriateness of evaluation measures: context.

Criterion	Measure	Operational definition	Mean (SD)
Information ethics	**Presence of ethics guidelines**	**Whether a digital library has guidelines for ethical issues**	**Scholar: 5.64 (1.062)** **Librarian: 5.00 (1.177)** **Overall: 5.33 (1.156)**
	Components of guidelines for ethics	Types of components in the guidelines for ethics	Scholar: 5.11 (1.370) Librarian: 4.85 (1.120) Overall: 4.98 (1.251)
Organizational mission	Conformity to organizational mission	To what extent a digital library conforms to organizational mission	Scholar: 5.52 (1.014) Librarian: 5.56 (1.050) Overall: 5.54 (1.023)
	Ways of support	In what ways a digital library supports organizational mission	Scholar: 5.56 (0.934) Librarian: 5.67 (1.000) Overall: 5.61 (0.960)
Targeted user community	User community engagement	Types of user community engagements	Scholar: 5.74 (0.903) Librarian: 5.56 (1.050) Overall: 5.65 (0.974)
	Level of user community engagement	To what extent a digital library engages in user communities	Scholar: 5.63 (0.926) Librarian: 5.59 (0.931) Overall: 5.61 (0.920)
Content sharing	Types of content sharing	Types of digital library content-sharing partners	Scholar: 5.00 (1.265) Librarian: 5.41 (0.747) Overall: 5.21 (1.044)
	Types of resources shared	Types of digital library items shared with partners	Scholar: 4.93 (1.299) Librarian: 5.35 (0.797) Overall: 5.13 (1.093)
	Number of items shared	Number of items shared with partners	Scholar: 5.12 (1.107) Librarian: 4.96 (0.898) Overall: 5.04 (0.999)
Collaboration	Types of collaboration	Types of digital library collaboration partners or stakeholders	Scholar: 4.81 (1.241) Librarian: 5.33 (0.832) Overall: 5.07 (1.079)
	Number of collaborations	Number of digital library collaborations in a specific period of time	Scholar: 4.76 (1.200) Librarian: 5.04 (0.898) Overall: 4.90 (1.053)
Social impact	Types of social impact	Types of social impact of a digital library on the society	Scholar: 5.44 (1.188) Librarian: 5.11 (0.801) Overall: 5.28 (1.017)
	Level of social impact	**To what extent a digital library influences society**	**Scholar: 5.69 (0.736)** **Librarian: 4.78 (0.892)** **Overall: 5.23 (0.933)**
Multilingual access	Types of languages	Types of languages supported by a digital library	Scholar: 5.37 (1.006) Librarian: 5.22 (1.050) Overall: 5.30 (1.021)

The distributions in the two groups differ significantly ($U = 262.5$, $p < .05$). The median responses in the scholar group and the librarian group for the measure of "level of social impact" are 6 and 5, respectively. The distributions in the two groups differ significantly ($U = 165$, $p < .05$). Overall, the ratings in this dimension were relatively lower than the ratings in the other dimensions. There was no measure that was rated over 6 on average. Most of the measures in this dimension are qualitative, which involves a more subjective assessment. It turned out that "number of collaborations" and "components of guidelines for ethics" were considered lower than "somewhat appropriate" by the participants.

Discussion

This study extends the research on digital library evaluation by presenting a comprehensive set of measures and ranking of their appropriateness for the associated criteria by digital library scholars and librarians. The two groups were asked to rank 174 measures for 82 criteria within 10 dimensions. Prior research on digital library evaluation

dimensions and criteria has found a significant divergence in how different stakeholders perceive the importance of criteria (Xie et al., 2018; Zhang, 2010). The findings from the first-round survey conducted by this research team indicated some consensus between scholars and academic librarians, but also noticeable differences in ranking the importance of criteria in dimensions such as "user engagement" and "preservation" (Xie et al., 2018).

However, the findings from the second-round survey, reported in this article, demonstrate a remarkable consensus between digital library scholars and practitioners on the appropriateness of measures associated with digital library evaluation criteria. Overall, the results support the hypothesis posed for this study that there is no significant difference in rating the appropriateness of each measure associated with a digital library evaluation criterion between the two groups. The Mann–Whitney test results fail to reject the null hypothesis for all measures in the dimensions of "collections," "information organization," "services," and "preservation," and the rejection in the other dimensions is from one or two measures. In the pool of 174 measures, this study found a lack of consensus for only 9 measures. Digital library scholars and librarians tended to agree more than disagree in their assessments of whether the proposed measures were appropriate for the evaluation criteria.

The two groups agree on a high rating (over 6) of the appropriateness of two or more measures in 7 out of the 10 dimensions: "collections," "information organization," "user interface," "systems/technology," "services," "preservation," and "administration." The similarities are particularly notable in the "collections" dimension for measures such as "compliance with digitization standards" and "quality specifications." In the "information organization" dimension, the high ranking of "compliance to the metadata standards" and "compliance to interoperability standards" indicates that measuring the quality of digital objects and their potential for resource discovery in a large-scale digital library environment is critical in digital library evaluation. The findings confirm the importance of measuring the quality of content and metadata reported in other studies (Park, 2009; Yan et al., 2014). Moreover, digital library scholars and librarians agree on several measures in the "interface design" dimension, where five measures were rated over 6 for their appropriateness. The high ranking of measures for the user interface criteria is reflective of the significant prior research in this area and well-established measures (Jeng, 2005; Kelly, 2014; Zhang, 2010).

The differences, although minor, are indicative of different perceptions and priorities among digital library scholars and library practitioners. The Mann–Whitney test results reject at least one null hypothesis for the rating of the appropriateness of measures associated with digital library evaluation criteria in multiple dimensions, including "interface design," "system and technology," "effects on users," "user engagement," "administration," and "context." In the "interface design" dimension, this study found a significant difference in rating the appropriateness of the measure of "design consistency," with librarians assigning a lower rating. While "design consistency" is an important criterion for digital library evaluation, librarians are aware that measuring it is difficult, especially for out-of-the-box solutions that have limited capabilities for the customization and control of design consistency.

Additionally, there are significant differences in the ranking of measures in the dimensions of "effects on users" and "context" that focus on evaluating the impact of digital libraries on research and teaching. These are important but challenging areas in digital library evaluation. Interestingly, the librarians ranked three measures higher in the "effects on users" dimension: "research productivity," "digital library uses for research," and "effects of digital library uses on teaching effectiveness." The scholars' lower ranking of these measures is somewhat surprising but perhaps reveals that faculty are more aware of the challenges in measuring the impact of digital libraries on research productivity and teaching. On the other hand, the scholars ranked the appropriateness of the "level of social impact" measure higher in the "context" dimension. This measure is intended to assess to what extent a digital library influences society, which can also be challenging to apply.

This study has practical implications for digital library evaluation by providing an extensive list of measures and ranking of their appropriateness by scholars and practitioners. The number of proposed criteria and measures can be overwhelming for researchers and librarians who want to conduct an evaluation study. In practice, it is not feasible or even advisable to conduct an evaluation study using 82 criteria and 174 measures. The value of this study for the field of practice is in assessing the appropriateness of the proposed measures. The data gathered in this study on the dimensions and most appropriate measures from the perspectives of librarians and scholars provides a foundation for planning effective evaluation of digital libraries using a list of agreed-on measures. Using data from this study, digital library evaluators can choose a smaller set of criteria and the measures ranked as the most appropriate to focus on selected aspects of digital libraries. On the other hand, the measures in three dimensions—"effects of users," "user engagement," and "context"—were not ranked highly for their appropriateness, which may be an indication of the challenges involved in evaluating using dimensions, and point to a need for more research in this area. Despite the comprehensiveness of the criteria and measures, this study has some limitations, especially in the measurement of use and reuse. As the field of digital library evaluation evolves, new approaches to evaluating

the reuse of digital objects are emerging, including altmetrics, reverse image lookup, and other reuse measures (Kelly, 2014; O'Gara et al., 2018; Reilly and Thompson, 2017; Yang and Dawson, 2018).

Conclusion

This study contributes to research and practice in digital library evaluation by presenting a set of measures for conducting evaluation studies and examining if the proposed measures are appropriate. Furthermore, this study strengthens the assessment of the measures by integrating the perspectives of digital library scholars and librarians. The list is comprehensive and goes beyond the measures commonly used for evaluating user interfaces and usability. In fact, the measures related to the quality of digital objects and metadata were ranked quite highly in this study. In contrast to research on evaluation criteria that demonstrates the divergence in the perceptions of multiple stakeholders, this study found significant agreement in identifying relevant measures and their ranking by scholars and librarians. The high ranking of measures related to interface design and the consensus between librarians and scholars demonstrate the strengths of digital library evaluation research in this area. The differences in ranking of some measures, although not major, tend to be in newer areas and reveal the different priorities and experiences of library practitioners and scholars. This study has practical implications for digital library evaluation and development, as it provides a list of criteria and measures to guide the practical evaluation of digital libraries in an academic setting. It may advance the field of digital library evaluation by assisting evaluators in selecting relevant and consistent measures, and enabling evaluation studies across different digital library platforms and organizations. Further research needs to apply relevant measures associated with selected evaluation criteria to assess specific digital libraries. The digital library measures can be further enhanced through actual digital library evaluation.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Iris Xie https://orcid.org/0000-0001-5901-4894
Soohyung Joo https://orcid.org/0000-0001-6770-5118

Note

1. See https://www.usnews.com/rankings

References

Albertson D (2015) Synthesizing visual digital library research to formulate a user-centered evaluation framework. *New Library World* 116(3–4): 122–135.

Candela L, Castelli D, Pagano P, et al. (2007) Setting the foundations of digital libraries: The DELOS manifesto. *D-Lib Magazine* 13(3/4). DOI: 10.1045/march2007-castelli.

Chowdhury G (2014) Sustainability of digital libraries: A conceptual model and a research framework. *International Journal on Digital Libraries* 14(3–4): 181–195.

Ferati M and Beyene WM (2017) Developing heuristics for evaluating the accessibility of digital library interfaces. In: Antona M and Stephanidis C (eds) *Universal Access in Human–Computer Interaction: Design and Development Approaches and Methods*. Cham: Springer, pp. 171–181.

Fuhr N, Hansen P, Mabe M, et al. (2001) Digital libraries: A generic classification and evaluation scheme. In: Constantopoulos P and Sølvberg IT (eds) *Research and Advanced Technology for Digital Libraries*. Berlin: Springer, pp. 187–199.

Fuhr N, Tsakonas G, Aalberg T, et al. (2007) Evaluation of digital libraries. *International Journal on Digital Libraries* 8(1): 21–38.

Gkoumas G and Lazarinis F (2015) Evaluation and usage scenarios of open source digital library and collection management tools. *Program* 49(3): 226–241.

Gonçalves MA, Fox EA and Watson LT (2008) Towards a digital library theory: A formal digital library ontology. *International Journal on Digital Libraries* 8(2): 91–114.

Gonçalves MA, Fox EA, Watson LT, et al. (2004) Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries. *ACM Transactions on Information Systems* 22(2): 270–312.

Gonçalves MA, Moreira BL, Fox EA, et al. (2007) "What is a good digital library?" A quality model for digital libraries. *Information Processing and Management* 43(5): 1416–1437.

Inal Y (2018) University students' heuristic usability inspection of the national library of Turkey website. *Aslib Journal of Information Management* 70(1): 66–77.

Huang S, Han Z, Yang B, et al. (2019) Factor identification and computation in the assessment of information security risks for digital libraries. *Journal of Librarianship and Information Science* 51(1): 78–94.

Jeng J (2005) Usability assessment of academic digital libraries: Effectiveness, efficiency, satisfaction, and learnability. *Libri* 55(2–3): 96–121.

Joo S and Xie I (2013) Evaluation constructs and criteria for digital libraries: A document analysis. In: Cool C and Ng K-B (eds) *Recent Developments in the Design, Construction, and Evaluation of Digital Libraries: Case Studies*. Hershey, PA: IGI Global, pp. 126–140.

Kelly EJ (2014) Assessment of digitized library and archives materials: A literature review. *Journal of Web Librarianship* 8(4): 384–403.

Kelly EJ (2018) Use of Louisiana's digital cultural heritage by Wikipedians. *Journal of Web Librarianship* 12(2): 85–106.

Kelly EJ, Muglia C, O'Gara G, et al. (2018) Measuring reuse of digital objects: Preliminary findings from the IMLS-funded

project. In: *JCDL'18: Proceedings of the 18th ACM/IEEE Joint Conference on Digital Libraries* (eds Chen J, Gonçalves MA, Allen JM, et al.), pp. 351–352. New York: Association for Computing Machinery.

Lai CF, Chiu PS, Huang YM, et al. (2014) An evaluation model for digital libraries' user interfaces using fuzzy AHP. *Electronic Library* 32(1): 83–95.

Li Y and Liu C (2019) Information resource, interface, and tasks as user interaction components for digital library evaluation. *Information Processing and Management* 56(3): 704–720.

Matusiak KK, Harper A and Heinbach C (2019) Use and reuse of visual resources in student papers and presentations. *Electronic Library* 37(3): 490–505.

Mune C and Agee A (2016) Are e-books for everyone? An evaluation of academic e-book platforms' accessibility features. *Journal of Electronic Resources Librarianship* 28(3): 172–182.

Noh Y (2010) A study on developing evaluation criteria for electronic resources in evaluation indicators of libraries. *Journal of Academic Librarianship* 36(1): 41–52.

Noh Y (2011) Evaluation of the resource utilization efficiency of university libraries using DEA techniques and a proposal of alternative evaluation variables. *Library Hi Tech* 29(4): 697–724.

Noh Y (2012) A study measuring the performance of electronic resources in academic libraries. *Aslib Proceedings* 64(2): 134–153.

O'Gara GM, Woolcott L, Kelly EJ, et al. (2018) Barriers and solutions to assessing digital library reuse: Preliminary findings. *Performance Measurement and Metrics* 19(3): 130–141.

Park JR (2009) Metadata quality in digital repositories: A survey of the current state of the art. *Cataloging and Classification Quarterly* 47(3–4): 213–228.

Reilly M and Thompson S (2014) Understanding ultimate use data and its implication for digital library management: A case study. *Journal of Web Librarianship* 8(2): 196–213.

Reilly M and Thompson S (2017) Reverse image lookup: Assessing digital library users and reuses. *Journal of Web Librarianship* 11(1): 56–68.

Saracevic T (2004) Evaluation of digital libraries. In: Agosti M and Führ N (eds) *Notes of the DELOS WP7 Workshop on the Evaluation of Digital Libraries*, Padua, Italy. Available at: http://dlib.ionio.gr/wp7/workshop2004_program.html

Stiller J and Petras V (2018) Learning from digital library evaluations: The European case. *ABI Technik* 38(1): 37–45.

Tsakonas G, Kapidakis S and Papatheodorou C (2004) Evaluation of user interaction in digital libraries. In: Agosti M and Führ N (eds) *Notes of the DELOS WP7 Workshop on the Evaluation of Digital Libraries*, Padua, Italy.

Xie H (2006) Evaluation of digital libraries: Criteria and problems from users' perspectives. *Library and Information Science Research* 28(3): 433–452.

Xie I, Joo S and Matusiak KK (2018) Multifaceted evaluation criteria of digital libraries in academic settings: Similarities and differences from different stakeholders. *Journal of Academic Librarianship* 44(6): 854–863.

Xie I and Matusiak KK (2016) *Discovery Digital Libraries: Theory and Practice*. Cambridge, MA: Elservier.

Yan Y, Zha X, Zhang J, et al. (2014) Comparing digital libraries with virtual communities from the perspective of e-quality. *Library Hi Tech* 32(1): 173–189.

Yang SQ and Dawson PH (2018) Altmetrics and their potential as an assessment tool for digital libraries. In: Kataria S, Anbu JPK, Gartner R and Sandhu G (eds) *2018 5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services (ETTLIS)*.: Institute of Electrical and Electronics Engineers, 351–354.

Zhang Y (2010) Developing a holistic model for digital library evaluation. *Journal of the American Society for Information Science and Technology* 61(1): 88–110.

Author biographies

Iris Xie is a professor in the School of Information Studies at the University of Wisconsin-Milwaukee. She received her PhD from the Rutgers University. Her research interests focus on the design and evaluation of digital libraries, interactive information retrieval, human-computer interaction, user studies and research methods. She has published two books: Interactive Information Retrieval in Digital Environments (IGI Publishing, 2008) and Discover Digital Libraries: Theory and Practice (Elsevier, 2016), co-authored with Krystyna Matusiak.

Soohyung Joo is an associate professor in the School of Information Science at the University of Kentucky. He received his PhD in Information Studies from the University of Wisconsin-Milwaukee. His areas of research interests include digital libraries, interactive information systems, data management, and applied data science.

Krystyna K. Matusiak is an associate professor in the Library & Information Science Program at the University of Denver. She received her Master's in Library and Information Science and Ph.D. from the University of Wisconsin-Milwaukee. Prior to accepting her position at the University of Denver, she worked as a digital collections librarian and served as a digitization consultant. Her research interests include digital libraries, digitization of cultural heritage, visual information, information-seeking behavior, and research methods.

Appendix 1. Mann–Whitney U tests: sum of scores, ranks, and statistics for the significant variables

Criterion: measure	Group	Mean rank	Sum of ranks	Test statistics
Consistency: design consistency	Scholars	32.27	903.50	$U = 230.50$
	Librarians	22.37	581.50	$p < .05$
Retrieval effectiveness: recall	Scholars	21.71	608.00	$U = 202.00$
	Librarians	34.52	932.00	$p < .05$
Research productivity: digital library uses for research	Scholars	23.67	686.50	$U = 251.50$
	Librarians	33.69	909.50	$p < .05$

Criterion: measure	Group	Mean rank	Sum of ranks	Test statistics
Instructional effectiveness: digital library uses in teaching	Scholars	22.89	641.00	$U = 235.50$
	Librarians	33.30	899.00	$p < .05$
Staffing: staff hours on digital libraries	Scholars	22.71	590.50	$U = 239.50$
	Librarians	31.13	840.50	$p < .05$
Staffing: student hours on digital libraries	Scholars	21.50	602.00	$U = 196.00$
	Librarians	34.74	938.00	$p < .05$
Site visit: unique page visits	Scholars	23.07	623.00	$U = 245.00$
	Librarians	31.93	862.00	$p < .05$
Information ethics: presence of ethics guidelines	Scholars	32.13	899.50	$U = 262.00$
	Librarians	23.72	640.50	$p < .05$
Social impact: level of social impact	Scholars	34.15	888.00	$U = 165.00$
	Librarians	20.11	543.00	$p < .05$