Library and Information Science Research 41 (2019) 78–84

Contents lists available at ScienceDirect

Library and Information Science Research

journal homepage: www.elsevier.com/locate/lisres

Development of a scale for data quality assessment in automated library systems

Mehri Shahbazi[a,*], Abdolhossein Farajpahlou[b], Farideh Osareh[b], Alireza Rahimi[c]

[a] Department of Knowledge and Information Science, Payame Noor University, Nakhi Street, Tehran, Iran.
[b] Department of Knowledge and Information Science, Shahid Chamran University, Ahvaz, Iran
[c] Medical Informatics Department, Faculty of Medical Management and Information Sciences, Isfahan University of Medical Sciences, Isfahan, Iran

ABSTRACT

A credible scale based on the opinions of system users was developed to evaluate and assess data quality in automated library systems (ALS). Development and testing were carried out in two stages. In the first stage, 77 dimensions for data quality which had been previously identified through a systematic literature review were used to develop scale items. The first draft of the scale was then distributed among a target population of ALS experts to solicit their opinions on the scale and the items. In the second stage, a revised version of the scale was distributed among the main study population, which included end users of the target systems. This stage used factor analysis to determine the final draft of the scale, which consists of 4 factors and 62 items. The 4 factors were named after the qualities of their associated items: Data Content Quality, Data Organizational Quality, Data Presentation Quality, and Data Usage Quality. This scale can help system managers identify and resolve potential problems in the systems they manage and can also aid in evaluating the quality of data sources based on the opinions of end users.

1. Introduction

Developers of information systems are usually concerned about end user satisfaction, however, there are few studies available which develop evaluation criteria for information systems based on user satisfaction. Evaluation criteria usually address concerns in categories such as management, technical issues, usage, boundary issues, policy, and customer issues (Farajpahlou, 2002). An evaluation is basically a judgment of worth; the ability to evaluate the return on investment provides the basis on which to choose between alternatives. Comparison of evaluation results with external standards, in the light of existing institutional realities which may be relevant, offers a path to evaluating the future trajectory of a program or service and provides an objective basis for decision making (Sharma, 2007).

The main goal of library-related information systems, whether database, website, portal, or automated library system (ALS), is to allow for search and retrieval from what have become enormous volumes of data. Since the success of data retrieval rests mainly on the soundness of the retrieval path, solving information retrieval path problems is one of the major concerns for developers of such systems and can lead to consumption of large amounts of time and financial resource for the individuals and organizations involved. However, despite the expenditure of much effort and money, end users of these systems still experience problems with data retrieval and quality. This study focuses on the different dimensions of data quality (DQ) and their significance according to end users.

In the information science literature the terms data, information, and even knowledge are often used interchangeably, which can lead to confusion. However, the literature that focuses on evaluation of quality generally agrees that dimensions of data and information quality are similar, and so in the present study the same stance is assumed. Also, according to Smart (2002) and H. Chen (2009) there is no standard, uniform and universal definition for DQ; however, the concept of "fitness for use" (Wang, 1998) has widely accepted for some time, and represents the spirit of the present study.

2. Problem statement

As stated earlier, satisfaction of end users is one of the most important concerns of information system designers and manufacturers, and it remains an elusive goal. One aspect of the problem is data quality. In many ALSs, search results do not match searchers' requests and expectations. Evidence suggests this inconsistency is rooted in the quality of data entered into the system (e.g., Dalcin, 2005; Fadli, 2013). Attention to, and assessment and evaluation of, DQ can be important in improving information retrieval in information systems and saving time for end users. However, to solve problems with regard to DQ in any organization or system and to bring about improvements, it is necessary to understand the dimensions of DQ.

No previous studies have attempted to assess DQ in ALSs, taken in the present study to mean computer-based library systems that manage input, processing, and bibliographic output. To reduce DQ problems in these systems it would be useful to have a scale which could assess DQ along different dimensions of ALSs. Given that the point is to facilitate

* Corresponding author.
E-mail address: shahbazi-mehri@pnu.ac.ir (M. Shahbazi).

https://doi.org/10.1016/j.lisr.2019.02.005
Received 22 April 2018; Received in revised form 25 September 2018; Accepted 8 February 2019
Available online 16 February 2019
0740-8188/

success and satisfaction for end users, it would seem obvious that their opinions should be taken into account in devising a scale.

Such a scale would benefit librarians, designers, and developers of ALSs. Not only might it elucidate the attitudes and opinions of different users about different dimensions of DQ in ALS, but it would also serve to identify potential system problems and promote improvement.

The following questions guided the present study:

RQ1. What are the most important dimensions of DQ in ALSs according to experts?

RQ2. What are the most important dimensions of DQ in ALSs according to end users?

RQ3. What would be the factors and items of a credible scale for assessing DQ in ALSs according to end users?

3. Previous research

Studies of different dimensions of DQ, especially in information systems, have gained increased attention. “Dimensions” are the factors that can help evaluate the quality of data. An extensive content analysis of DQ studies by Shahbazi (2017) identified 77 dimensions of DQ referenced in the literature, though there are admittedly some fine shades of meaning between some dimensions (Table 1). Other studies have attempted to propose models, methods, or scales for assessing and measuring the quality of data or information in various information systems (Rahimi, Farajpahlou, Osareh, & Shahbazi, 2017). A selected list of such studies can be seen in Table 2.

Most studies have considered primarily the structure, services, and presentation of models for assessing ALSs and similar systems. Some also have examined customer satisfaction, but most of these only address one or a few aspects of DQ as part of a larger list of aspects (Bhardwaj & Shukla, 2000; Joint, 2006; Osaniyi, 2010; Ramesh, 1998; Sharma, 2007; Taole, 2008). No studies have specifically attempted to assess DQ in ALSs.

4. Methodology

4.1. Scale development

This research used both bibliometric and survey methods, applied in two stages. First, using resources from the literature in the field of data quality, various dimensions of data quality were extracted and used to create items for a questionnaire, which in this first stage consisted of 77 dimensions and 147 items. The items in the scale were informative statements created by the researchers based on the definition of each dimension of DQ. The factors were sets of different items which emerged after categorization using exploratory factor analysis and were used as subscales to assess DQ in connection with end users. The validity of the questionnaire was determined using content validity at each stage by consulting experts and faculty members in the fields of informatics, scientometrics, and computer sciences. For this purpose, at each step, after refining each item, the list of factors and relevant items was prepared and was given to these experts to resolve grammatical and conceptual issues that could be spotted in the content. Experts' views were then applied in revising the survey. The modified version of the scale which was then used in the second stage of the study had 70 dimensions and 127 items. This scale formed the basis for a final scale that was developed on the basis of the opinions of end users about DQ in ALSs.

Cronbach's alpha was used to determine the reliability of the questionnaire. Usually the reliability of a questionnaire is considered acceptable if this coefficient is higher than 0.75 (Dayani, 2005). Cronbach's alpha calculated in each stage of the study showed the questionnaire to be highly reliable in both stages (Table 3).

The main tool used for data collection was a Likert-type scale that included a number of factors and items. The population for the final stage consisted of two groups: end users of the ALSs in libraries both in Isfahan and Ahwaz in Iran, as well as systems experts in the same libraries, i.e., managers of ALS and those involved in design and development of the systems. End users included both library customers and employees.

Table 1
Dimensions of DQ.

Accessibility	Definition	Readability
Accuracy	Density	Recoverability
Adaptability	Documentation	Redundancy
Age	Duplication	Relevance
Applicability	ease of manipulation	Reliability
Appropriate amount of data	Ease of operation	Representational consistency
Attractiveness	Ease of use	Response time
Attribute granularity	Effectiveness	Robustness
Availability	Efficiency	Security
Believability	Expiration	Semantic consistency
Clarity	Flexibility	Source's information
Coherence	Flexibility	Specialization
Comparability	Freshness	Stability
Compatibility	Homogeneity	Storage capability
Completeness	Identifiability	Structural consistency
Complexity	Informativeness	Sufficiency
Comprehensiveness	Interactivity	Time stability
Concise representation	Interoperability	Timeliness
Confidentiality	Interpretability	Traceability
Conformity	Meaningfulness	Unambiguousness
Convenience	Naturalness	Uncertainty
Correctness	Novelty	Understandability
Credibility	Objectivity	Uniqueness
Currency	Obtainability	Usability
Customer support	Organization	Value-adding
Data volume	Precision	

4.2. Populations

The academic libraries of Isfahan and Ahwaz included libraries in public (Payam-e-Noor University, the University of Applied Science and Technology) and nonprofit universities supervised by the ministries of Science, Research and Technology, and Health, Treatment and Medical Education as well as libraries in Islamic Azad University branches. These libraries were included because they use the most common ALSs available in Persian. They were also more accessible for the researchers.

The population in the first stage consisted of experts in the fields of scientometrics and informatics. Given the small number of members in this population, sampling was not necessary and the census method was adopted. All managers of libraries and information centers, all librarians with experience with ALSs in Isfahan and Ahwaz, and all experts in the firms developing ALSs made up a total of 90 individuals, of whom 79 agreed to participate in the study. The majority of participants (78.5%) had a degree in librarianship or informatics. Most participants held either a bachelor's (53.2%) or master's (29.1%) degree; only 3.8% held a PhD degree, and 6.4% held either high school or associate degrees.

The population for the second stage consisted of 120,849 potential users of ALSs in academic libraries of Isfahan and Ahwaz, that is, all the academic members in various departments and faculties in the five universities. Since exploratory factor analysis was used in this stage to modify the questionnaire and prepare the final scale, the sampling ratio of 1 over 5 was used based on the number of items against the relevant population (Beshlideh, 2012). Hence, given the 127 items in the questionnaire, against 120,849 potential individuals for the study, 635 potential end users were selected using stratified random sampling; all then completed the questionnaire. Among these participants, 57.9% had bachelor's degrees, 21.9% held a master's, 16.7% held PhDs, and 3.5% had associate degrees. Fields of study included humanities (34.5%), medical sciences (26.9%), technology and engineering (25.7%), sciences (7.7%), and arts (5.2%).

Table 2
Selected data quality studies.

No.	Study focus	Source
1	Websites or portals	Jeong & Lambert, 2001; Chun Chung Joshua, 2006; Herrera-Viedma et al., 2007; Caro et al., 2008; Calero et al., 2008; Leite, Gonçalves, Teixeira, & Rocha, 2015
2	Relational databases	Parssian, 2003
3	Cooperative information systems	Scannapieco, Virgillito, Marchetti, Mecella, & Baldoni, 2004
4	Genomic storage	Martinez, 2007
5	Fuzzy neural networks	Xiaojuan et al., 2008
6	Enterprise resource planning (ERP)	Xiaosong, Zhen, Meng, Dainuan, & Ting, 2008; Haug et al., 2009
7	Electronic commerce systems	H. Chen, 2009
8	Electronic education systems	Alkhattabi, Neagu, & Cullen, 2011
9	RFID systems	Bardaki, Kourouthanassis, & Pramatari, 2011; Togt, Bakker, & Jaspers, 2011
10	Software requirements for web applications	Guerra-García et al., 2013
11	Monitoring center quality management system	Bergvall, Lindquist, & Norén, 2014
12	Hospital information systems	Ratnaningtyas & Surendro, 2013; Rahimi, Liaw, Ray, Taggart, & Yu, 2014; Liaw, Taggart, Yu, & Rahimi, 2014; Rahimi et al., 2014
13	Information systems in general	C. Chen, 2002; Stvilia, 2006

Table 3
Cronbach's α for the scale at each stage.

Stage	α
First	0.983
Second	0.979

5. Findings

5.1. Most important dimensions according to experts

The main goal of the first stage was to categorize different dimensions of DQ based on the opinions of the experts who were involved with the ALSs. The questionnaire at this stage consisted of 147 items (using Likert-like scales) that were categorized in 77 dimensions. Representational consistency, with an average score of 3.71, proved to be the most important dimension of DQ in experts' views. The next 10 important dimensions were documentation, structural consistency, compatibility, conformity, response time, confidentiality, storage capability, semantic consistency, homogeneity, and comparability.

The principal purpose of this stage was to derive a modified questionnaire with high reliability for application in the second stage of the study. This was achieved by using correlation testing, which resulted in removing some unimportant dimensions among which low correlations existed. This method also identified items whose elimination would lead to increased correlation and reliability (through Cronbach's alpha).

Data analysis in the first stage showed that eliminating 21 items in 7 dimensions of unambiguousness, customer support, interactivity, ease of manipulation, correctness, redundancy, and recoverability actually increased the value of Cronbach's alpha for the questionnaire. Therefore, these items and dimensions were eliminated and a modified version of the questionnaire was developed with 127 items in 70 dimensions, and was administered in the second stage.

5.2. Most important dimensions according to end users

The second stage of the study sought to identify the most important dimensions of DQ based on the views of end users. The modified questionnaire resulting from the first stage was distributed to the 635 potential end users.

In this stage, exploratory factor analysis was used to allow the researchers to categorize the scales and observed variables without applying a precondition to the data. After gathering the necessary data, principal component analysis (PCA) was used to analyze the results. Fitness of data for factor analysis was investigated using anti-image correlation, Kaiser–Meyer–Olkin testing, and Bartlett's Test of Sphericity (Beshlideh, 2012; Pallant, 2010). From the correlation indexes that were higher than 0.3 and calculation of the anti-image correlation matrix, it was concluded that sample size was sufficient for factor analysis. Also, all diagonal elements of this matrix were greater than 0.50 and between 0.70 and 0.99. The Kaiser–Meyer–Olkin value (0.941) also showed the credibility of data and a sufficient sample size for factor analysis (Table 4). Further, the calculated value for Bartlett's Test of Sphericity (0.000) showed that data had reached statistical significance level and factor analysis on the correlation matrix was possible. If Bartlett's test of sphericity is lower than 0.05, this shows that the component matrix is not an identity matrix, and adequate correlation exists for factor analysis. Also, the determinant of the data matrix (6.35) is higher than 0.00001, indicating a lack of multiple linearity problems, which confirms that the data were suitable for PCA.

The communalities table in this analysis showed that the largest factor loadings belonged to understandability, traceability, availability, novelty and freshness. In the view of end users, these items and their dimensions are the most important items and dimensions of DQ in ALSs. The lowest factor loadings belonged to security, duplication, concise representation and credibility (Table 5).

5.3. Final scale

In order to determine the items and dimensions for the final scale, PCA and extraction of primary factors were carried out using eigenvalues and the Scree test (Beshlideh, 2012; Gildeh & Moradi, 2012; Pallant, 2010). Eigenvalues and total variance explained for factors were calculated. The first three columns of Table 6 show factors, their eigenvalues, total variance explained, and cumulative percentage. The other columns of this table show total explained variance for four main factors before and after rotation. Eigenvalues showed that 32 factors had variance larger than 1 (only the top 10 factors are presented in this table). The explained variance of the first four factors is the highest before and after rotation and these four factors together can explain 34% of total explained variance.

A scree diagram (Fig. 1) also shows the eigenvalue for each factors. The slope of the curve toward the x-axis changes between factor 3 and 4. This figure shows that the first factor is a dominating factor. Based on

Table 4
Methods for investigating the suitability of data for factor analysis.

KMO and Bartlett's test		
Determinan		6.35
Kaiser-Meyer-Olkin measure of sampling adequacy		0.941
Bartlett's test of sphericity	approximate χ^2	38,491.393
	df	8001
	sig.	0.000

Table 5
Highest and lowest load factors.

Item	Factor loading	
All records are understandable for users	523	Highest factor loading
It is easy to track the necessary resources using their accession cataloging numbers	493	
Current situation (availability) of resources is shown to users	490	
The data are related to novel and new ideas	483	
Data are novel and new	473	
Access to library resources is possible according to the predefined user level	120	Lowest factor loading
Access to data is limited for every user	127	
It is possible to input duplicate data if necessary	155	
The data expected to be presented for a document are shown all together	166	
The data demonstrate the credibility of the author of each resource in the library	177	

the eigenvalues and the scree diagram, the first 4 factors were extracted for further analysis. The researchers, considering the survey items and the factor loading, explored the proper categorization for the factors and the survey items by several proper rotations. The component matrix shows that pure variables with load factors of 0.3 or higher only happen in one factor. Therefore, in order to better interpret and categorize factors and items in a factor, rotation was required. Hence, at this stage data analysis began with a Varimax rotation. In categorization of factors and items in this rotation (Table 7), some items had two or more factors with factor loading higher than 0.3. Furthermore, the component matrix also showed a correlation higher than 0.3 among the factors. This allows application of the Oblimin rotation in the analysis process (Pallant, 2010) and so the component matrix was repeated using Oblimin rotation (Table 8).

The Oblimin rotation was carried out once with a rotation of 0.33 and then again with a rotation of 0.4 The values resulting from this rotation and the factor categorization matrixes were analyzed by researchers and several experts and finally an Oblimin rotation of 0.4 was accepted as the proper rotation at this stage. This rotation led to categorization of the items into 4 factors and 62 items (Table 9). Thus the final DQ scale for assessing data quality in the ALSs according to end users was developed.

Each of the 4 factors form a subscale. In this stage, 24 dimensions of quality were eliminated: concise representation, obtainability, security, believability, complexity, duplication, compatibility, structural consistency, semantic consistency, attractiveness, currency, organization, adaptability, ease of operation, precision, naturalness, applicability, reliability, identifiability, homogeneity, definition, density, age, and uncertainty.

The final stage of the scale creation usually includes naming of the factors or subscales that turn up in the final rotation. To this end, researchers used the opinions of professors and faculty members as well as considering the nature of each subscale to create the factor names seen in Table 9. The subscales (factors) were accordingly named Data Content Quality (17 items), Data Organizational Quality (6 items named), Data Presentation Quality (12 items), and Data Usage Quality (27 items).

6. Discussion

Interestingly, the opinions of experts were different from those of end users about the most important dimensions of DQ. Experts felt that the most important dimensions mostly related to structure and content of data in the target systems, while end users believed that dimensions related to data retrieval as well as the novelty of data in the systems were important. Furthermore, dimensions of compatibility, structural consistency, semantic consistency and homogeneity, which were some of the most important dimensions in the views of the experts, were found to be the least important according to end users. The final four categories or subscales of DQ, derived from extensive analysis, seem to make intuitive sense, as they are concerned with content quality, quality of content organization, quality of data display and organization, and data utility.

More attention to the dimensions eliminated after analysis of end user data also revealed that these dimensions fit in categories that are less concerned with quality and therefore less likely to be used for evaluation of DQ by end users. The list of the least important dimensions according to experts and end users complied with findings reported in over a decade of studies, including Najjar (2002); Missier et al. (2003); Rajamani (2006); Stvilia (2006); Martinez (2007); Herrera-Viedma, Peis, Morales-del-Castillo, Alonso, and Anaya (2007); Xiaojuan, Shurong, Zhaolin, and Peng (2008); Calero, Caro, and Piattini (2008); Caro, Calero, Caballero, and Piattini (2008); Haug, Arlbjørn, and Pedersen (2009); Michel-Verkerke (2012); Saberi and Mohd (2013); Guerra-García, Caballero, and Piattini (2013). Some of these studies studied DQ dimensions in their own fields, however, unlike the current study, none of them focused on end user views.

The items included in the scale proposed in this study also pinpoint some problems with DQ in information systems. These are problems mentioned by researchers in the literature but no study has attempted a comprehensive categorization of these problems. The results of these studies, based on their views of the problems and related dimensions,

Table 6
Total variance explained and eigenvalues.

Factors	Eigenvalues			Total explained variance for 4 factors before rotation			Total explained variance for 4 factors after rotation		
	Total	Variance percent	Cumulative percent	Total	Variance percent	Cumulative percent	Total	Variance percent	Cumulative percent
1	34.056	26.816	26.816	34.056	26.816	26.816	12.723	10.018	10.018
2	4.156	3.272	30.088	4.156	3.272	30.088	10.524	8.287	18.305
3	3.069	2.417	32.505	3.069	2.417	32.505	10.471	8.245	26.55
4	2.539	1.999	34.504	2.539	1.999	34.504	10.101	7.954	34.504
5	2.223	1.751	36.254						
6	2.133	1.680	37.934						
7	1.877	1.478	39.412						
8	1.792	1411	40.823						
9	1.747	1.375	42.198						
10	1.68	1.323	43.521						

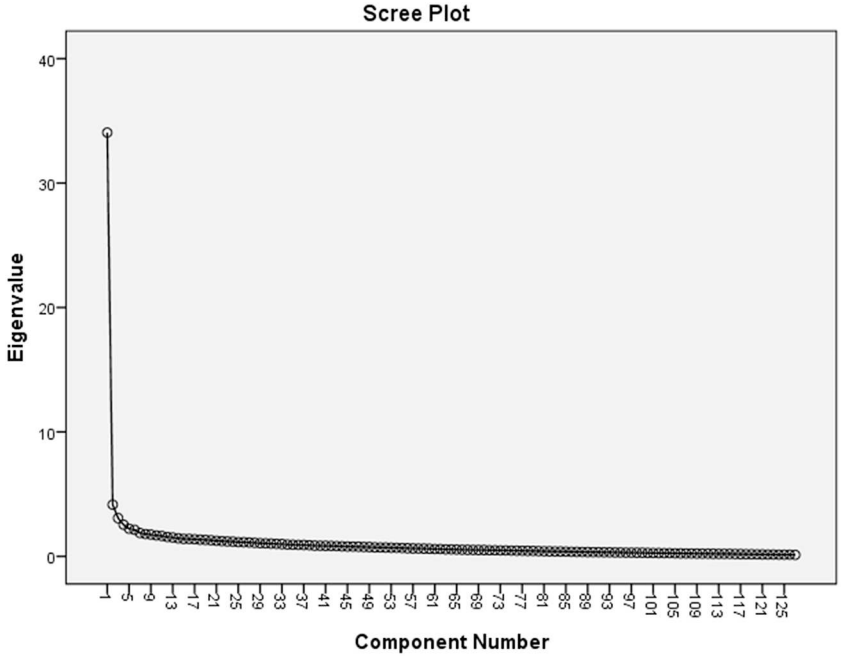

Fig. 1. Scree curve of components.

Table 7
Component correlation in Varimax rotation.

Component transformation matrix				
Factor	1	2	3	4
1	0.557	0.497	0.471	0.471
2	−0.450	0.223	0.741	−0.445
3	−0.415	−0.493	0.315	0.697
4	−0.561	0.678	−0.360	0.308

Notes: Extraction method: principal component analysis; rotation method: varimax with Kaiser normalization.

Table 8
Component matrix in Oblimin rotation.

Component correlation matrix				
Factor	1	2	3	4
1	1.000	0.415	0.374	−0.518
2	0.415	1.000	0.251	−0.322
3	0.374	0.251	1.000	−0.438
4	−0.518	−0.322	−0.438	1.000

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.
Notes: Extraction method: principal component analysis; rotation method: oblimin with Kaiser normalization.

are comparable to the items designed in the proposed scale in the present study. Strong, Lee, and Wang (1997) divided problems with DQ into three categories:

- intrinsic problems, such as lack of match between similar data from different sources;
- accessibility problems, for example lack of access to computerized data (often due to low system resources), time and effort spent to gain access to information, limited processing capacity for image and text data, and high volume or the low processing speed; and
- data usage problems, such as the existence of incomplete data as well as contradictory displays of data which leads to weakening of the quality of text data. This includes some mal-presentations of text that are ambiguous either in content or in form, while important in essence as values. Such data could potentially mislead or be misunderstood by users.

These problems are directly and indirectly represented in the design of the items of DQ evaluation mentioned in the present study.

Stvilia (2006) focused on a number of problems with data usage, including accessibility, accuracy, authority, cohesiveness, complexity, consistency, informativeness, naturalness, relevancy, and verifiability. He believes that these problems derive from complexity, language ambiguity, poor structure, typographical errors, the use of alternative words, confusion or software clutter, lack of supportive sources, lack of accurate resource review, bias in review of resources, multiple perspectives and unbalanced coverage of different perspectives, lack of details, low readability, use of different terms for similar concepts, use of different structures and styles for particular types of data, mismatch between recommended styles and guides, differences in cultural or linguistic meanings, standards, content redundancy, non-fluent text, unrelated or out of context content, lack of references to original sources, lack of access to original sources, and instability caused by editing sabotage (p. 160). The problems mentioned by Stvilia are found similarly among the items of the proposed scale in the present study.

Berti-Équille (2007) discusses the problem of the existence of similar records with different titles and believes that this can be related to such procedures as reconciliation, refinement or unification, subject matching, duplicate removal, citation matching, identifying uncertainties, identifying nature, and separation by nature. He also emphasizes problems related to obsolescence and outdated data (p. 192). The important thing to note when comparing the results of current study with those of Berti-Équille is his emphasis on duplication, which turned out to be considered not important by participants, hence it was eliminated from the final scale.

6.1. Limitations

Some limitations of the present research resulted from the research design. Some items were eliminated in stage 1, before end users saw the scale, and so end users did not have the opportunity to present their views on those items. Therefore for those items there could be no

Table 9
Item categorization in 4 factors using oblimin rotation.

No	Factor	Items
1	1st = Data Content Quality	When users need information, available data can meet this need using newest resources.
2		Data are related to new and novel ideas
3		User knows the age of data points
4		Retrieved data meet the research needs of users.
5		Retrieved data are exactly what users need based on their goal and applications
6		Available data covers all subject areas searched by users
7		Search results always match what user wants
8		Data is presented in a way that can easily be transferred to other systems
9		Data available in the system can meet the demands
10		Data related to specific topics are regularly updated
11		Concepts are presented hierarchically and from whole to part
12		After searching desired keywords, retrieved data is enough for the user
13		Data present in retrieved records is correct and without error in matching the original
14		Users are able to correct search information
15		Data are new
16		Retrieved data present the users with all the needed information about resources
17		Users are able to search in any desired field
18	2nd = Data Organizational Quality	Data is sorted according to principles of cataloging
19		Data is organized enough to reduce the time needed by users
20		Subject field correctly shows resources' subject
21		Available data is useful for reaching the intended resource
22		All books' information in considered in search
23		Data obviously follow a defined standard
24	3rd = Data Presentation Quality	Data is accessible through sharing or purchased credits
25		No incorrect or unintelligible words are present in data
26		There is data on library hosting the resources
27		The credibility of each data can be determined using scientific evidence
28		There is an option near each retrieved result which shows the table of content or abstract of the resource in a separate page
29		There is information in the system on how to enter data
30		There is an option near each retrieved result which shows the bibliographic information of the resource in a separate page
31		The standards used help users access to other information in the system
32		When data are entered, all subsystems are accessible
33		There is data on resources' authors
34		The number of retrieved records relevant to search keywords is higher than the number of irrelevant results
35		Data are presented in detail
36	4th = Data Usage Quality	Using cataloging number, it is easy to find the resources
37		All retrieved records are understandable for users
38		The language used in this systems is understandable for users
39		All retrieved records belong to library resources and are credible
40		Symbols used in data display are intelligible
41		Data can easily be confirmed by checking resources
42		Retrieved data are useful
43		Available data are suitably used
44		Data related to resources are always easily accessible
45		Search results can be saved based on users' needs
46		When attempting to access unauthorized data, users receive an error massage
47		It is possible to save the data
48		All similar data (e.g. like information on all books) are shown similarly
49		The availability status of resources is shown to users
50		Data are new and novel
51		Data available in the system can easily be interpreted
52		Data retrieved for each resource cover all of its aspects
53		Data retrieved for each record is unique
54		Data in each record is fully reliable. For example, cataloging number shows the exact location of books
55		Data coding s similar and comparable to other systems
56		Data are easy to use
57		Data have suitable readability
58		System can retrieve searched information in seconds
59		Retrieved data matches real data
60		Data lack any complexity or difficulty
61		Data in each record is enough to know the resource
62		Data are cost-effective

comparison between experts and end users. Also, some dimensions were conceptual or abstract and therefore more open to personal interpretation than other more concrete items.

7. Conclusion

This research has led to the construction of a scale that can be used to evaluate the quality of data in ALSs. The proposed scale is unique in its method of DQ assessment. The main and the most important benefits of this scale lie in the help it can provide to library managers and system staff in identifying potential problems in ALSs and resolving these problems, especially when they affect user satisfaction. The dimensions and items covered in the proposed scale make it a useful tool for evaluating DQ in ALSs, especially where views of end users are desired. Furthermore, along with the content analysis by Shahbazi (2017), the fairly comprehensive coverage of items and their organization into meaningful categories would be useful for anyone studying the evaluation of data quality in information settings.

References

Alkhattabi, M., Neagu, D., & Cullen, A. (2011). Assessing information quality of e-learning systems: A web mining approach. *Computers in Human Behavior, 27*, 862–873.

Bardaki, C., Kourouthanassis, P., & Pramatari, K. (2011). Modeling the information completeness of object tracking systems. *The Journal of Strategic Information Systems, 20*(3), 268–282.

Bergvall, T., Lindquist, M., & Norén, G. N. (2014). vigiGrade: A tool to identify well-documented individual case reports and highlight systematic data quality issues. *Drug Safety, 37*, 65–77.

Berti-Équille, L. (2007). Data quality awareness: A case study for cost optimal association rule mining. *Knowledge and Information Systems, 11*(2), 191–215.

Beshlideh, K. (2012). *Research methodes and statistical analysis of research examples using SPSS and AMOS* (16th ed.). Ahwaz, Iran: Shahid Chamran University Press.

Bhardwaj, R. K., & Shukla, R. K. (2000). A practical approach to library automation. *Library Progress (International), 20*(1), 1–9.

Calero, C., Caro, A., & Piattini, M. (2008). An applicable data quality model for web portal data consumers. *World Wide Web, 11*, 465–484.

Caro, A., Calero, C., Caballero, I., & Piattini, M. (2008). A proposal for a set of attributes relevant for web portal data quality. *Software Quality Journal, 16*, 513–542.

Chen, C. (2002). *A framework for optimizing data quality given limited resources (Unpublished doctoral dissertation)*. Tempe: Arizona State University.

Chen, H. (2009). *Modeling information quality in agent-based e-commerce systems (Unpublished master's dissertation)*. Canada: University of Ottawa.

Chun Chung Joshua, P. (2006). *On the use of the appropriateness and cohesiveness Web data quality dimensions for finding high quality Web pages (Unpublished doctoral dissertation)*. Kowloon: Hong Kong University of Science and Technology.

Dalcin, E. C. (2005). *Data quality concepts and techniques applied to taxonomic databases (Unpublished doctoral dissertation)*. England: University of Southampton.

Dayani, M. H. (2005). *Methods of research in librarianship*. Tehran, Iran: Computer Library.

Fadli, M. S. (2013). Critical success factors and data quality in accounting information systems in Indonesian cooperative enterprises: An empirical examination. *Interdisciplinary Journal of Contemporary Research in Business, 5*, 321–338.

Farajpahlou, A. H. (2002). Criteria for the success of automated library systems: Iranian experience (application and test of the related scale). *Library Review, 51*, 364–372.

Gildeh, B. S., & Moradi, V. (2012). Fuzzy tolerance region and process capability analysis. In B. Y. Cao, & X. J. Xioe (Eds.). *Fuzzy engineering and operations research* (pp. 183–193). Berlin, Germany: Springer.

Guerra-García, C., Caballero, I., & Piattini, M. (2013). Capturing data quality requirements for web applications by means of DQ_WebRE. *Information Systems Frontiers, 15*, 433–445.

Haug, A., Arlbjørn, J. S., & Pedersen, A. (2009). A classification model of ERP system data quality. *Industrial Management & Data Systems, 109*, 1053–1069.

Herrera-Viedma, E., Peis, E., Morales-del-Castillo, J. M., Alonso, S., & Anaya, K. (2007). A fuzzy linguistic model to evaluate the quality of web sites that store XML documents. *International Journal of Approximate Reasoning, 46*, 226–253.

Jeong, M., & Lambert, C. U. (2001). Adaptation of an information quality framework to measure customers' behavioral intentions to use lodging web sites. *International Journal of Hospitality Management, 20*(2), 129–146.

Joint, N. (2006). Evaluating library software and its fitness for purpose. *Library Review, 55*, 393–402.

Leite, P., Gonçalves, J., Teixeira, P., & Rocha, Á. (2015). A model for the evaluation of data quality in health unit websites. *Health Informatics Journal, 22*, 479–495. https://doi.org/10.1177/1460458214567003.

Liaw, S. T., Taggart, J., Yu, H., & Rahimi, A. (2014). Electronic health records and disease registries to support integrated care in a health neighbourhood: An ontology-based methodology. *AMIA Joint Summits on Translational Science: Proceedings, 2014*, 50–54. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4419761/.

Martinez, A. (2007). *BIODQ: A model for data quality estimation and management in biological databases (Unpublished doctoral dissertation)*. Gainesville: University of Florida.

Michel-Verkerke, M. B. (2012). Information quality of a nursing information system depends on the nurses: A combined quantitative and qualitative evaluation. *International Journal of Medical Informatics, 81*, 662–673.

Missier, P., Lalk, G., Verykios, V., Grillo, F., Lorusso, T., & Angeletti, P. (2003). Improving data quality in practice: A case study in the Italian public administration. *Distributed and Parallel Databases, 13*, 135–160.

Najjar, L. (2002). *The impact of information quality and ergonomics on service quality in the banking industry (Unpublished doctoral dissertation)*. Lincoln: University of Nebraska.

Osaniyi, L. (2010). Evaluating the X-lib library automation system at Babcock university, Nigeria: A case study. *Information Development, 26*(1), 87–97. https://doi.org/10.1177/0266666909358306.

Pallant, J. (2010). *Analysis of behavioral sciences data using SPSS* (3 ed.). Tehran, Iran: Frozesh (Persian version, A. Rezaie, translator).

Parssian, A. H. (2003). *Assessing information quality for relational databases (Unpubished doctoral dissertation)*. Dallas: University of Texas.

Rahimi, A., Farajpahlou, H., Osareh, F., & Shahbazi, M. (2017). Developments of research in evaluation of data and information quality in information systems since the year 2000. *Iranian Journal of Information Processing and Management, 33*(2), 915–944.

Rahimi, A., Liaw, S. T., Ray, P., Taggart, J., & Yu, H. (2014). Ontological specification of quality of chronic disease data in EHRs to support decision analytics: A realist review. *Decision Analytics, 1*(5), 1–31.

Rahimi, A., Parameswaran, N., Ray, P. K., Taggart, J., Yu, H., & Liaw, S. T. (2014). Development of a methodological approach for data quality ontology in diabetes management. *International Journal of E-health and Medical Communuications, 5*(3), 58–77.

Rajamani, S. (2006). *Information systems with emphasis on business rules for demographics: A Kaizen strategy for data quality in public health informatics (Unpubished doctoral dissertation)*. Minneapolis: University of Minesota.

Ramesh, L. S. R. C. V. (1998). Technical problems in university libraries on automation: An overview. *Herald of Library Science, 37*(3–4), 165–173.

Ratnaningtyas, D. D., & Surendro, K. (2013). Information quality improvement model on hospital information system using six sigma. *Procedia Technology, 9*, 1166–1172.

Saberi, S., & Mohd, M. (2013). Is data quality an influential factor on web portals' visibility? *Procedia Technology, 11*, 834–839.

Scannapieco, M., Virgillito, A., Marchetti, C., Mecella, M., & Baldoni, R. (2004). The DaQuinCIS architecture: A platform for exchanging and improving data quality in cooperative information system. *Information Systems Frontiers, 29*, 551–582.

Shahbazi, M. (2017). *Developing and validating a scale for assessment of data quality in computerized library systems based on experts and end-users viewpoints (Unpublished doctoral dissertation)*. Ahvaz, Iran: Shahid Chamran University.

Sharma, S. (2007). *Library automation software packages used in academic libraries of Nepal: A comparative study. (Unpublished associate's degree dissertation)*. New Delhi, India: Council of Scientific and Industrial Research.

Smart, K. L. (2002). Assessing quality documents. *ACM Journal of Computer Documentation, 26*(3), 130–140.

Strong, D., Lee, Y., & Wang, R. (1997). Data quality in context. *Communications of the ACM, 40*(5), 103–110.

Stvilia, B. (2006). *Measuring information quality (Unpublished doctoral dissertation)*. University of Illinois at Urbana-Champaign.

Taole, N. (2008). *Evaluation of the Innopac library system in selected consortia and libraries in the southern African region: Implications for the Lesotho Library Consortium (Unpublished doctoral dissertation)*. South Africa: University Of Pretoria.

Togt, R., Bakker, P. J. M., & Jaspers, M. W. M. (2011). A framework for performance and data quality assessment of radio frequency identification (RFID) systems in health care settings. *Journal of Biomedical Informatics, 44*, 372–383.

Wang, R. Y. (1998). A product perspective on total data qualtiy. *Communications of the ACM, 41*(2), 58–65.

Xiaojuan, B., Shurong, N., Zhaolin, X., & Peng, C. (2008). Novel method for the evaluation of data quality based on fuzzy control. *Journal of Systems Engineering and Electronics, 19*, 606–610.

Xiaosong, Z., Zhen, H., Meng, Z., Dainuan, Y., & Ting, Z. (2008). The application study of ERP data quality assessment and improvement methodology. *IEEE Conference on Industrial Electronics and Applications, 3*, 1036–1039.

Mehri Shahbazi is a faculty member in the Department of Library and Information Science, Payame Noor University, Isfahan, Iran. She holds a PhD and MS in library and information studies from Shahid Chamran University, Ahvaz, Iran. Her research interests include information systems, data quality, knowledge management, information organization, library systems, and children's librarianship. Her articles have appeared in Iranian and international journals and several of her monographs have been published by Payame Noor University.

Abdolhossein Farajpahlou is a professor in the Department of Knowledge and Information Science, Shahid Chamran University, Ahvaz, Iran, where his work focuses on library management as well as quality management and information technology applications in library and information centers. He is also the head of the planning committee for library and information science education at the Ministry of Science, Research and Technology in Iran. He received his PhD from the University of New South Wales, Sydney, Australia, and has published in Iranian and international journals.

Farideh Osareh is a professor in the Department of Knowledge and Information Science, Shahid Chamran University, Ahvaz, Iran, She received her PhD from the University of New South Wales, Sydney, Australia. She has published and translated 10 books and has written more than 120 articles for journals in Farsi and English. Her main research interest include scientometrics, co-authorship, social networks, and information visualization.

Alireza Rahimi is an assistant professor and medical informatics specialist at Isfahan University of Medical Sciences, Iran. He earned his PhD in health and medical informatics studies from the University of New South Wales, Australia, where he received the outstanding doctoral thesis award for his research on developing and testing an ontologically based algorithm to accurately identify patients with Type 2 diabetes mellitus. He founded the Health Information Technology Research Center at Isfahan University of Medical Sciences. His research interests include data quality in clinical information systems, design of ontological based approaches, IT human resource management, information organization, and automated library systems. He has published in journals such as *Advanced Biomedical Research*, *International Journal of E-Health and Medical Communications*, and *International Journal of Medical Informatics*, among others, and authored a chapter in "E-Health and Telemedicine: Concepts, Methodologies, Tools and Applications" (IGI Global, 2016).