# Information Quality Framework for e-Learning Systems

Mona Alkhattabi*

School of Computing, Informatics and Media
University of Bradford
Bradford, BD7 1DP, UK
E-mail: monaalkhattabi@gmail.com

Daniel Neagu

School of Computing, Informatics and Media
University of Bradford
Bradford, BD7 1DP, UK
E-mail: D.Neagu@bradford.ac.uk

Andrea Cullen

School of Computing, Informatics and Media
University of Bradford
Bradford, BD7 1DP, UK
E-mail: A.J.Cullen@bradford.ac.uk

*Corresponding author

**Abstract:** Information quality frameworks are developed to measure the quality of information systems, generally from the designers' viewpoint. The recent proliferation of e-services and e-learning particularly raises the need for a new quality framework in the context of e-learning systems. This paper proposes a new information quality framework, with 14 information quality attributes grouped in three quality dimensions: intrinsic, contextual representation and accessibility. This framework could be useful to e-learning systems designers, providers and users as it provides a comprehensive indication of the quality of information in such systems. We report results based on original questionnaire data and factor analysis supporting our conclusions.

**Keywords:** E-learning; Information quality; Information systems; quality frameworks.

**Biographical notes**: Dr. Mona Alkhattabi graduated the PhD studies in Computing from the School of Computing, Informatics and Media of the University of Bradford in 2010. She is currently an Assistant Professor in the College of Computer Science at Imam Mohammad bin Saud University, Saudi Arabia. Mona has over 5 years educational experience in the field of Mathematics, with a bachelor degree from Education College in Riyadh, Saudi Arabia. After she was awarded an MSc in Computing with Distinction from the University of Bradford in 2006, she started her research in the quality of e-learning content. More details are included in her personal webpage: http://www.malkhattabi.com.

Dr. Daniel Neagu is senior lecturer in computing with the University of Bradford, where he was appointed in 2002. His research focuses on Artificial Intelligence techniques applied to Data Quality, Visual Arts, Toxicology, Software Engineering, OODbs and Web Semantics. Processing information to identify key attributes and making unknown, hidden or distributed information accessible to people plays a vital role in the progress of science and technology, thus Daniel's work addresses applications as diverse as e-learning, health, engineering, and entertainment. More details are included in his webpage: http://www.inf.brad.ac.uk/staff/index.php?type=p&u=dneagu.

Dr. Andrea Cullen has many years industrial experience as a computer programmer and systems analyst working on projects throughout the UK. Her academic career spans two disciplines: computer science; and management. She is currently a senior lecturer in computing at the University of Bradford, teaching operations management, business systems security and e-business. Main areas of research include: e-commerce, e-government, and issues associated with IS security within organisations. More details are included in her webpage: http://www.inf.brad.ac.uk/staff/index.php?type=p&u=ajcullen.

---

## 1. Introduction

Today quality is considered a crucial issue for education in general, and for e-learning in particular. Currently there are two recognized challenges in e-learning: the demand for overall interoperability and the request for high quality. Moreover, quality cannot be expressed and set by a simple definition, since in itself quality is a very abstract notion. The specified context and the perspectives of users need to be taken into account when defining quality in e-learning. It is also essential to classify suitable criteria to address quality (Stracke, 2006).

In the literature, there is a wide interest in information quality provided by information systems in general. However taking into account that quality on the web is a complex concept and its measurement is expected to be multidimensional in nature (Aladwani & Palvia, 2002), the prime issue in evaluating the quality of any information system is identifying the criteria by which the quality is determined (Buyukozkan, Ruan, & Feyzioglu, 2007). The criteria are a result of the multidimensional and interdependent nature of quality in information systems, and are dependent on the objectives and the context of the system.

This paper is part of a wider research project aiming to define metrics to determine the quality of the content provided by distributed learning materials, for integrating intelligent agent technologies as a means of gathering information for quality evaluation.

This paper focuses on concepts of information quality in the context of e-learning systems, particularly on identifying the key dimensions for information quality from the users' perspective in order to build a quality framework to measure the quality of the content provided by e-learning systems. It is essential to identify quality dimensions accurately as they provide the building blocks for further research into the quality of e-learning information systems in general. Great attention has been given to ensure the accuracy of the diminutions defined in this paper. In our study, Wang & Strong's data quality framework (Wang & Strong, 1996) was extended and used as the reference point owing to its popularity and acceptance by the information systems quality community.

The rest of the paper is organized as follows. The next section reconsiders the meaning of e-learning and its definition including the concept of quality in e-learning systems. Section three reviews previous research related to information systems quality frameworks and proposes the first draft of the new framework. In section 4, we discuss our work to collect learners' opinion to identify information quality characteristics in e-learning systems and the preliminary results. Data analysis and the revised framework format are presented in sections 5 and 6 respectively, followed by the conclusion and future work in the final section.

## 2. E-learning

The term e-learning is used in literature and commercial applications to describe many fields such as online learning, web-based training, distance learning, distributed learning, virtual learning, or technology-based training. During the last decade, e-learning was defined in literature in different ways. In general, most definitions for e-learning are used to express the exploitation of the technologies which can be used to deliver learning (or learning materials) in an electronic format, most likely via the internet (Gerhard & Mayr, 2002 ). Within the same line of defining e-learning as the delivery of the content through the technical channels, Paulsen more generally describes online learning as "the use of a computer network to present or distribute some educational content" (Paulsen, 2002). Psaromiligkos and Retalis consider e-learning systems as those which utilize the internet as a delivery medium for static learning resources, such as instructional files, or as an interface into interactive content (Psaromiligkos & Retalis, 2003).

The previous definitions look at e-learning in general. In more detail, e-learning can be seen in the form of courses or in the form of modules [separate parts of course's objects] and smaller learning materials. In addition, e-learning can include synchronous or asynchronous interaction.

Considering that there are two main types of e-learning: asynchronous and synchronous, depending on the interaction between learner and teacher, we will now discuss these in more detail. Synchronous e-learning environments require tutors and learners or the online classmates to be online at the same time, where live interactions take place between them. However, the focus of our research will be on the case where students are logging into and using the system independently of other students and staff members. This fits firmly into the general definition of the asynchronous e-learning environment. In this context, Doherty describes an Asynchronous Learning Network [ALN] as a variety of e-learning systems which distribute learning materials and concepts in one direction at a time (Doherty, 1998). Moreover, Spencer and Hiltz express ALN as a place where learners can interact with learning materials, tutors and other learner/s through the internet at different times and from different places (Spencer & Hiltz, 2001).

The position adopted in this research is that e-learning covers the technology used to distribute the learning materials, the quality of these materials, and the interaction with learners. We adopt in the definition of e-learning used in this paper these dimensions as described by the European Commission in (Gerhard & Mayr, 2002 p.2):

> "*the use of new multimedia technologies and the internet to improve the quality of learning by facilitating access to resources and services as well as remote exchange and collaborations*"

## 2.1. The concept of quality in e-learning systems

The definition of e-learning adopted in this research represents three fundamental dimensions: technology, access and quality. However, the focus in our study will be on quality, which is considered a crucial issue for education in general, and for e-learning in particular. Currently there are two recognized challenges in e-learning: the demand for overall interoperability and the request for [high] quality. Moreover, quality cannot be expressed and set by a simple definition, since in itself quality is a very abstract notion. The specified context and the perspectives of users need to be taken into account when defining quality in e-learning. It is also essential to classify suitable criteria to address this quality (Stracke, 2006).

Although it is important to set standards for information quality, this is a difficult and complex issue because there is no formal definition of information quality, as quality is dependent on the criteria applied to it. Furthermore, it is dependent on the targets, the environment and from which viewpoint we look at the information quality, that is, from the provider or the consumer perspective.

This section of the paper will discuss concepts of quality in e-learning generally. Despite efforts to reach a comprehensive, universal definition of quality in e-learning, there is still a fundamental ambiguity surrounding it and we will approach this further in the paper's conclusions.

One position is to consider quality as an evaluation of excellence, a stance that is primarily adopted by universities and education institutions. For example, in universities, quality teaching and learning are promoted as the top priority, giving less attention to criteria or measurements regarding teaching input into courses, the learning outcomes, and the interactivity with the system (Crisp, 2002). Another trend is to consider the improvement in quality, by moving beyond the set of conceptions in the direction of flexible processes of negotiation which needs a very high level of quality capability from those involved (Ehlers, Goertz, Hildebrandt, & Pawlowski, 2005).

Quality can be viewed and considered from different aspects. In this context the SunTrust Equitable report (Close, Humphreys, & Ruttenbur, 2000) illustrates what they perceive to be the value chain in e-learning in the form of a pyramid. Content is the most critical factor of e-learning as it forms the base of the value pyramid. In fact, to be able to use the internet as a tool to improve learning, the content should not distract learners, but increase their interest for learning. Learning tools and enablers are also important in the learning procedure. In reality, providers of learning platforms and knowledge management systems are key factors in the successful delivery of content. The providers need infrastructure to deliver learning content. Moreover, learning service providers [LSP] are the distribution channels for content providers. One of the challenges which face these knowledge hubs and LSPs is to ensure that the learners are receiving fresh content. Companies focused on educational e-tailing [electronic retailing] are completing the value pyramid of e-learning.

From their e-learning value pyramid it can be observed that content is the most critical component of learning through the internet. In a similar manner, we will find that the measurement of the quality of content delivered by e-learning is the most important criteria and the most influential in the overall level of learning quality.

## 3. Information quality frameworks

Although it is important to set standards for information quality, it is a difficult and complex issue particularly in the area of information systems because there is no formal definition of information quality, as quality is dependent on the criteria applied to it. Furthermore, it is dependent on the targets, the environment and from which viewpoint we look at the information quality, that is, from the provider or the consumer perspective. Moreover information quality is both a task-dependent and a subjective concept; Juran summarises these aspects of quality in his quality definition as "fitness for use" (Juran., 1974).

However, it is common to define information quality on the internet by identifying the main dimensions of the quality. For that purpose information quality frameworks are widely used to identify the important quality dimensions as described by Porter (Porter, 1991).

During the last years, much work has been done [as will be discussed later in this section] to build quality frameworks for information quality dimensions. In the past, research in information quality frameworks focused on data quality, but due to the recent development of internet technologies, information systems today are providing users information, not only data. Therefore, research attention shifted to focus on information quality frameworks. However, still in some studies the term "information quality" is interchangeable with "data quality". Discussion on this issue is outside the scope of this paper, but we will return to it in the future work.

This part of the paper focuses on the Wang & Strong's data quality framework and reviews quality models, which were published since. We also present our proposed framework, which will be a result of the expansion of the original model to support identifying the key dimensions for information quality in e-learning systems.

### 3.1. Wang and Strong data quality framework

Wang & Strong's data quality framework, one of the most comprehensive, popular, remarkable and cited data quality framework, was established by Richard Wang and Diana Strong in 1996 (Wang & Strong, 1996). Their framework was designed empirically by asking users to give their viewpoint about the relevance of the information quality dimensions to capture the most important aspects of data quality to the data consumer. Lately, several quality management projects in business and government have successfully used this framework. Their hierarchical conceptual framework of data quality is shown in Figure 1[1].

In their framework, Wang and Strong classified quality dimensions into four groups (Wang & Strong, 1996):

- ***Intrinsic data quality:*** refers to the quality dimensions originated from the data in its own. This aspect of quality is independent of the user's perspective and context.

[1] Reproduced from (Wang & Strong, 1996) by kind permission of the author.

- ***Contextual data quality:*** focuses on the aspect of information quality within the context of the task at hand. In this group, the quality dimensions are subjective preferences of the user. Contrary to the first group, data quality dimensions cannot be assessed without considering the users viewpoint about their use of provided information.
- ***Representational data quality:*** is related to the representation of information within the systems.
- ***Accessibility data quality:*** refers to the quality aspects concerned into accessing distributed information.

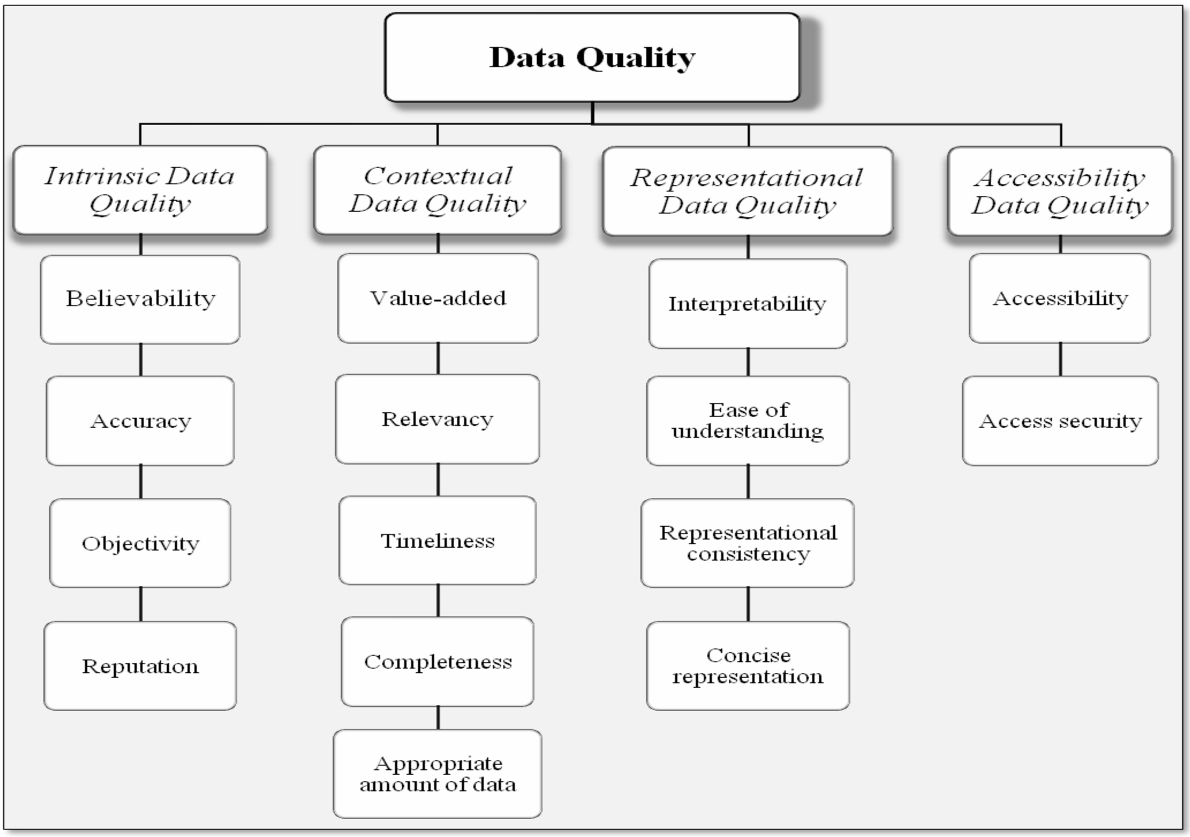


**Figure 1.** ***Wang & Strong's data quality framework***

Although their quality model will provide a good basis for our research to measure information quality in e-learning systems along the dimensions of this framework, it should be extended to include any undiscovered quality dimensions that occurred in the lately published research in the area of the quality in information systems.

## 3.2. Information quality in recent years

After Wang & Strong’s data quality framework, diverse research efforts were spent in order to identify information quality dimensions in different contexts as shown in Table 1.

We extended Wang & Strong's data quality framework by examining seventeen frameworks within the recently published literature. In general, we found that there are nineteen quality dimensions permanently used in most of the frameworks. Fifteen of them are already used in Wang & Strong framework. Table 1 summarises the occurrence of these dimensions within the examined frameworks. Table 2 gives the frequency of the appearances for every dimension along the examined frameworks.

These dimensions are grouped into four categories as defined within the Wang & Strong's framework. The nineteen initial quality dimensions, which were identified in the examined frameworks, will be used as an extended framework and therefore as a fundamental base to discover the important quality dimensions from the users' perspective in the context of e-learning systems.

**Table 1. Comparison between the emergences of quality dimensions in different information quality framework (part 1)**

| Information quality | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Quality factors** | **Quality dimensions** | Gertz & Managing 1996 (Gertz, 1996) | Redman 1996 (Redman, 1996) | Zeist & Hendriks 1996 (Zeist & Hendriks, 1996) | Jarke & Vassiliou 1997 (Jarke & Vassiliou, 1997) | Chen et all 1998 (Chen, Zhu, & Wang, 1998) | Alexander & Tate 1999 (Alexander & Tate, 1999) |
| **Intrinsic dimensions** | Accuracy | √ | √ | √ | √ | √ | √ |
| | Believability | | | | √ | | √ |
| | Consistency | | | | √ | | |
| | Objectivity | | √ | | | | √ |
| | Reputation | | | | | | √ |
| **Contextual dimensions** | Appropriate amount of data | √ | √ | √ | | √ | √ |
| | Completeness | √ | √ | | √ | √ | |
| | Relevancy | | √ | √ | √ | √ | √ |
| | Timeliness | √ | | √ | √ | √ | √ |
| | Value-added | | | | | | |
| | Verifiability | | √ | | | | |
| **Representational dimensions** | Concise representation | | √ | | | | |
| | Ease of understanding | | | √ | | | |
| | Interpretability | | √ | | √ | | |
| | Representational consistency | | √ | | √ | | √ |
| **Accessibility dimensions** | Accessibility | | | √ | | | √ |
| | Access Security | | | √ | √ | | |
| | Availability | √ | √ | | √ | | |
| | Response time | | | | √ | √ | |

**Table 1. Comparison between the emergences of quality dimensions in different information quality framework (Part 2)**

| | **Information quality frameworks** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Katerattanakul & Siau 1999 (Katerattanakul & Siau, 1999) | Shanks & Corbitt 1999 (Shanks & Corbitt, 1999) | Dedeke 2000 (Dedeke, 2000) | Zhu & Gauch 2000 (Zhu & Gauch, 2000) | Leung 2001 (Leung, 2001) | Eppler & Muenzenmayer 2002 (Eppler & Muenzenmayer , 2002) | Kahn et all 2002 (Kahn, Strong, & Wang, 2002) | Klein 2002 (Klein, 2002) | Mecell a 2002 (Mecel la et al., 2002) | Liu & Han 2005 (Liu & Han, 2005) | Besiki et all 2007 (Besiki, Gasser, Twidale, & Smith, 2007) |
| | √ | √ | √ | | √ | √ | | √ | √ | √ | √ |
| | √ | √ | | √ | | √ | √ | | | | |
| | | √ | √ | | | √ | √ | | √ | √ | √ |
| | | √ | | √ | √ | √ | √ | √ | | | |
| | √ | √ | | | | | | | | √ | |
| | | | √ | | √ | √ | √ | √ | | | |
| | | √ | √ | | √ | √ | √ | √ | √ | √ | √ |
| | √ | | √ | √ | √ | | | √ | | √ | √ |
| | | √ | √ | √ | √ | √ | √ | √ | √ | √ | |
| | | | | | | | | | | | |
| | | | | | | √ | | | | | √ |
| | | | | | | √ | | | | | |
| | | | √ | | | √ | √ | | | | |
| | | | | | | | | | | | |
| | √ | | | | | √ | √ | | | √ | |
| | √ | √ | √ | √ | √ | √ | √ | | | √ | √ |
| | | | | | √ | √ | | | | | √ |
| | | | | | | | | | | | |
| | | | √ | √ | √ | √ | √ | | | √ | |

## 3.3. The proposal for the extended framework

We propose to update Wang and Strong's data quality framework initially comprising another four quality dimensions. Therefore, the extending quality framework consists of four quality factors and nineteen quality dimensions as shown in Figure 2.

**Table 2. Dimensions' frequencies in the examined frameworks**

| Quality dimensions | Frequency |
|---|---|
| Accuracy | 15 |
| Believability | 7 |
| Consistency | 8 |
| Objectivity | 8 |
| Reputation | 4 |
| Appropriate amount of data | 10 |
| Completeness | 13 |
| Relevancy | 12 |
| Timeliness | 14 |
| Value-added | 0 |
| Verifiability | 3 |
| Concise representation | 2 |
| Ease of understanding | 4 |
| Interpretability | 2 |
| Representational consistency | 7 |
| Accessibility | 11 |
| Access Security | 5 |
| Availability | 3 |
| Response time | 8 |

## 4. The survey

Although quality frameworks help in the measurement procedure, defining the quality using a framework is not enough because as mentioned before information quality is dependent on the application context. For that reason the identified quality dimensions were arranged in a questionnaire format to determine the users' view of the relative importance of quality dimensions in an e-learning system. This questionnaire[1] seeks to gather the views of end-users about the importance of information quality dimensions in e-learning systems. It also gives an indication about the importance and relevancy of these quality dimensions for the users, which will help in ranking these dimensions in order to develop an information quality framework for quality metrics to measure the quality of information provided by e-learning systems.

This investigation was a cross-section survey performed on a sample chosen from a population of persons involved in academic work and dealing with e-learning systems in a regular basis. Respondents were included both of learners and teachers. The questionnaire was distributed to the respondents via e-mail because of its reduced cost, decrease short transfer time and its convenience for respondents. Surveymethods.com, an online survey software application, was used to create the survey, deploy it via e-mail, and collect and analyze respondent data through its graphical based analysis module. The

---

[1] The survey can be accessed from www.elearningquality.com

questionnaire was planned to take less than five minutes to complete. The questionnaire consisted of three parts:

**Part 1** gives a brief profile of the respondent.

**Part 2** addresses the user's attitude and usage of the internet in general and e-learning systems specifically.

**Part 3** asks respondents to rank the nineteen quality dimensions in order of their importance.

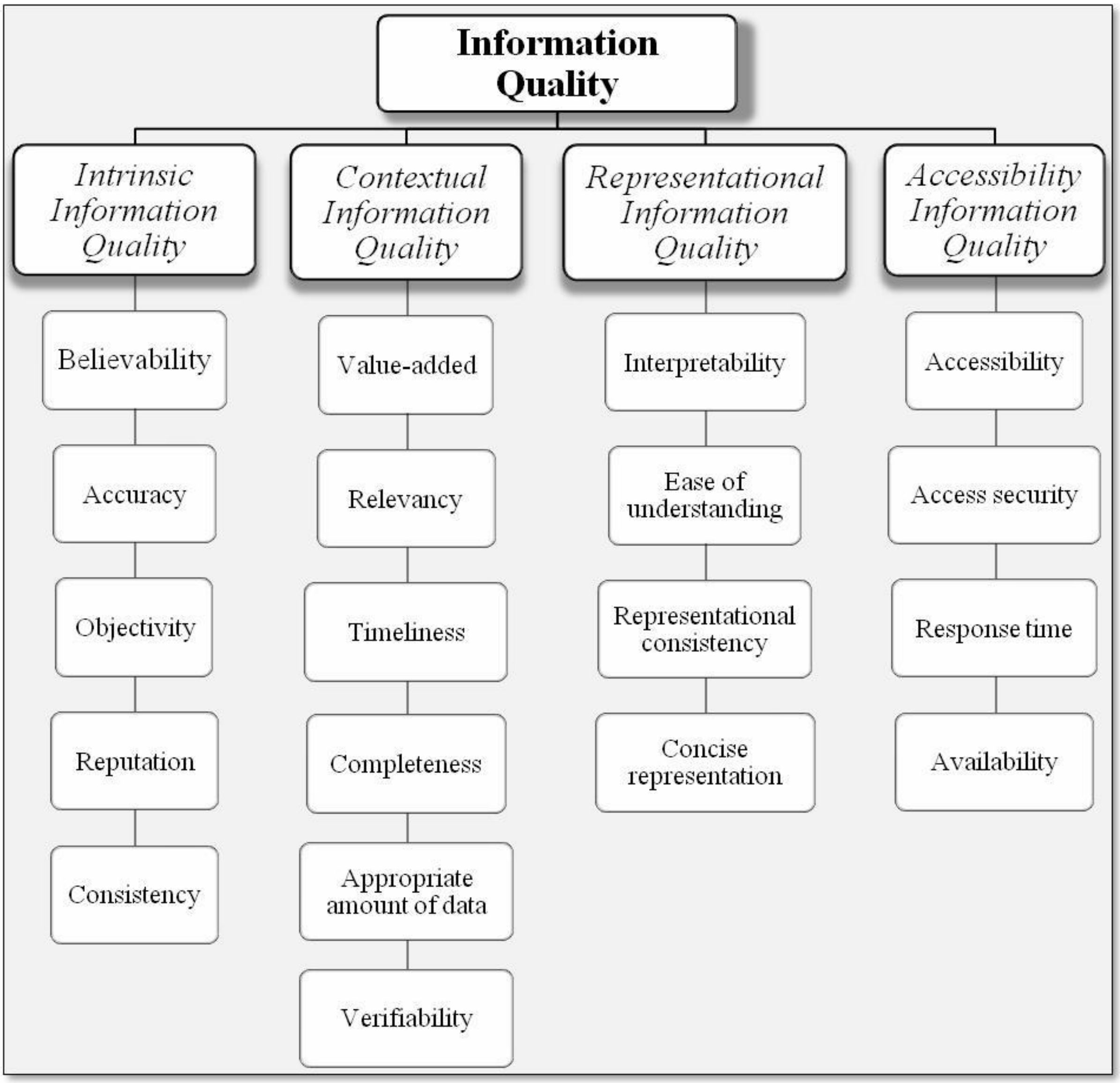


**Figure 2. The extended framework**

We collected responses from 315 e-learning system users[1], from 24 different countries, 46% of the respondents were from Saudi Arabia, 26% from United Kingdom, 12% from Romania and the rest of the respondents were from the 21 remaining countries. 57% of the participants were females, and 43% were males. All the respondents in the sample were e-learning users from different learning institutes. Of the respondents that contributed, the majority [66%] use e-learning as learners, and 29% as teachers and authors of the learning materials while 5% use e-learning systems for other purposes such as librarians and technicians. In addition, participants are holding various qualifications, 40% were holding Bachelor's degree, 33% have Master's degree, and 20% have PhD while the remaining 7% hold those listed as others.

## 5. Data analysis

We analyse the collected data from the third part of the questionnaire using SPSS to identify the most important quality dimensions in the area of e-learning systems and to build the final quality framework.

First, we conducted a frequency analysis for each variable to check for major mistakes and missing values. The results for variables frequency analysis in each dimension show that the data is valid and ready to be analysed.

Reliability is the level to which research results would be the same if the investigation was to be repeated with a different sample or at a later date. In this research, the most accepted test of inter-item consistency reliability is the Cronbach's coefficient alpha (L.J. Cronbach, 1951; L.J. Cronbach, 1971). Based on Sekaran reliabilities less than 0.6 are considered to be poor, those in the 0.7 range are acceptable, and those over 0.8 are good (Sekaran, 2000). The closer to 1.0 the better the reliability coefficient is. It is generally agreed that the minimum acceptable value of Cronbach's alpha is 0.70 (Pallant, 2005; Peter, 1979), but this could be reduced to 0.6 for exploratory research (Robinson, Shaver, & Wrightsman, 1991). The Cronbach's alpha values for the dimensions in each quality factor gave an acceptable reliability level with 0.712, 0.735, 0.781, and 0.625 for intrinsic, contextual, representational and accessibility information quality respectively.

Screening the data responding to Churchill's recommendation will increase the reliability levels (Churchill & Gilbert, 1979). So, the collected data was screened by discarding items that showed very small corrected item-total correlations [<0.40]. Because of this test, we delete timeliness and value-added variables from contextual factor, and access security variable from accessibility factor, which leaves only 16 dimensions in the framework. As a result, the reliability coefficient increased to 0.712, 0.748, 0.781, 0.668 for intrinsic, contextual, representational and accessibility factors respectively.

The next stage was conducting a factor analysis procedure with *varimax* rotation to check the dimensionality of the construct. To choose the cut-off value, there is no fixed measure. It depends on the purpose of the study on hand. Haire recommended that item loadings >0.30 are considered significant, >0.40 are more important, and >0.50 are considered very significant (Hair, Tatham, Anderson , & Black, 1998). While the aim of this study is to recognize the most important and significant quality attributes, we decided to use a cut-off point of 0.50 for item loadings and *eigenvalue* of 1.

[1] As recorded on 5th of March 2009

The determinant of the correlation matrix[1] is 0.002, which is greater than the necessary value of 0.00001. As a result, we are confident that multicollinearity will not cause any problems for our data (Field, 2000).

**Table 3. KMO and Bartlett's Test**

| | | |
|---|---|---|
| **Kaiser-Meyer-Olkin Measure of Sampling Adequacy.** | | .879 |
| **Bartlett's Test of Sphericity** | Approx. Chi-Square | 1845.750 |
| | Df | 120.000 |
| | Sig. | .000 |

**Table 4: Total Variance Explained**

| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 6.058 | 37.865 | 37.865 | 6.058 | 37.865 | 37.865 | 3.735 | 23.343 | 23.343 |
| 2 | 1.401 | 8.753 | 46.619 | 1.401 | 8.753 | 46.619 | 2.627 | 16.418 | 39.762 |
| 3 | 1.188 | 7.424 | 54.043 | 1.188 | 7.424 | 54.043 | 2.285 | 14.281 | 54.043 |
| 4 | .974 | 6.090 | 60.133 | | | | | | |
| 5 | .879 | 5.497 | 65.630 | | | | | | |
| 6 | .786 | 4.914 | 70.544 | | | | | | |
| 7 | .657 | 4.106 | 74.650 | | | | | | |
| 8 | .598 | 3.740 | 78.390 | | | | | | |
| 9 | .565 | 3.530 | 81.920 | | | | | | |
| 10 | .556 | 3.477 | 85.397 | | | | | | |
| 11 | .532 | 3.324 | 88.721 | | | | | | |
| 12 | .455 | 2.843 | 91.565 | | | | | | |
| 13 | .412 | 2.576 | 94.141 | | | | | | |
| 14 | .366 | 2.286 | 96.426 | | | | | | |
| 15 | .306 | 1.911 | 98.338 | | | | | | |
| 16 | .266 | 1.662 | 100.000 | | | | | | |

Extraction Method: Principal Component Analysis.

The Kaser-Meyer-Oklin [KMO] measure of sampling adequacy and Bartlett's test of sphericity are illustrated in Table 3. The KMO static is a value between 0 and 1. A value close to 1 indicates that patterns of correlation are fairly compact and as a result factor analysis should gives distinct and reliable factors (Field, 2000). Values between 0.5 and 0.7 are average, values between 0.7 and 0.8 are good, values between 0.8 and 0.9

[1] See appendix I

are great and values above 0.9 are excellent (Hutcheson & Sofroniou, 1999). Moreover, the significant value for the Bartlet's test should be less than 0.05 (Field, 2000). In our data, the value is .879, which is in the range of being great and the Bartlett's test is highly significant. Therefore, we are confident that factor analysis is appropriate for our data.

In addition to examining the overall KMO statistic, it is essential to check the diagonal elements of the anti-image correlation matrix[1] [which illustrates the KMO value for individual variables]; as in the overall KMO value these values have to be greater than 0.50 for all variables (Field, 2000). For our data, the values are in the range between [0.828 … 0.934].

Concerning the sample size, Comrey and Lee stated that 300 is a good sample size for factor analysis, 100 is poor while 1000 is excellent (Comrey & Lee, 1992). Since the number of our sample exceeds 300 respondents we should be confident that the sample size is appropriate for this type of analysis.

Table 4 lists the *eigenvalues* associated with each factor before extraction, after extraction and after rotation.

Before extraction, SPSS has identified 16 factors within the data set. SPSS then extracts all factors with *eigenvalues* greater than 1, which leave us with three factors.

From the *scree plot* shown in Fig 3, we can see that the point of inflexion on the curve on three factors which is in conformity with the results shown in Table 4. Thus, the most suitable way is to stick with three factors.

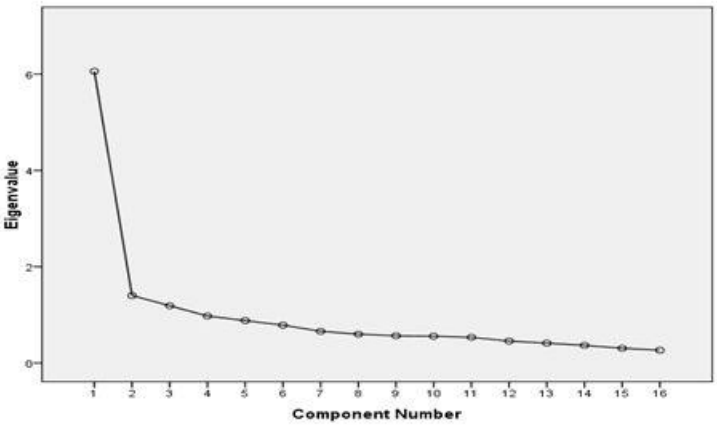


**Figure 3. Scree plot**

Table 5 shows the rotated component matrix, which is the matrix of the factor loadings for each variable onto each factor. Factors loading less than 0.5 have not been displayed because we asked for these lodgings to be suppressed. As a result, we discarded these suppressed variables, which are Consistency and Interpretability, which leave only 14 variables in total.

Analysis finding shows that there are three information quality factors in e-learning systems not four, as proposed previously. We recognized that contextual and

[1] See appendix II

representational quality factors are measuring the same aspects from e-learning systems users' perspective. Therefore, we propose a new framework, with 14 dimensions of information quality in e-learning systems to measure three quality factors: Intrinsic, Contextual representation and Accessibility information quality.

We then calculated Cronbach's alpha values for the variables in each new factor, which gives a good reliability level with 0.842, 0.697, and 0.665, for Intrinsic, Contextual representation and Accessibility information quality respectively. The new proposed framework is shown in Figure 4.

**Table 5. Rotated Component Matrix [a]**

| Quality attributes | Component | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| Believability | | | .689 |
| Accuracy | | | .736 |
| Objectivity | | | .765 |
| Reputation | .673 | | |
| Consistency | | | |
| Relevancy | | .541 | |
| Completeness | .582 | | |
| Amount Of Information | .607 | | |
| Verifiability | .695 | | |
| Interpretability | | | |
| Understandability | .643 | | |
| Representational Consistency | .596 | | |
| Conciseness | .809 | | |
| Accessibility | | .667 | |
| Response Time | | .623 | |
| Availability | | .782 | |

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 5 iterations.

Linear regression then was used to predict the factor scores from the variables. correlation coefficient, and can be obtained by squaring the "*part corr*" provided by SPSS "β in the equation bellow". For example for Completeness in the first factor, that is $0.156^2 = 2.434\%$. These statistics will sum to less than 100%. To get them to sum to 100%, we divided each by the sum of all. So we can calculate the relative importance for each variable in the correlated factor we can use the following equation:

$$relative\ importance\ for\ vi = \frac{\beta i^2}{\sum_i \beta i^2} \quad , i = 1, \ldots, 14,$$

where βi is the partial correlation for the variable $vi$ in the corresponding factor. The same logic was conducted to define the relative importance for each factor in the overall quality.

The zero-order correlations in Appendix3 are the loadings. One could define the relative importance of a variable as the amount by which the explained variance in the factor is reduced if the variable is removed from the regression model. That statistic measure is the squared semi-partial.

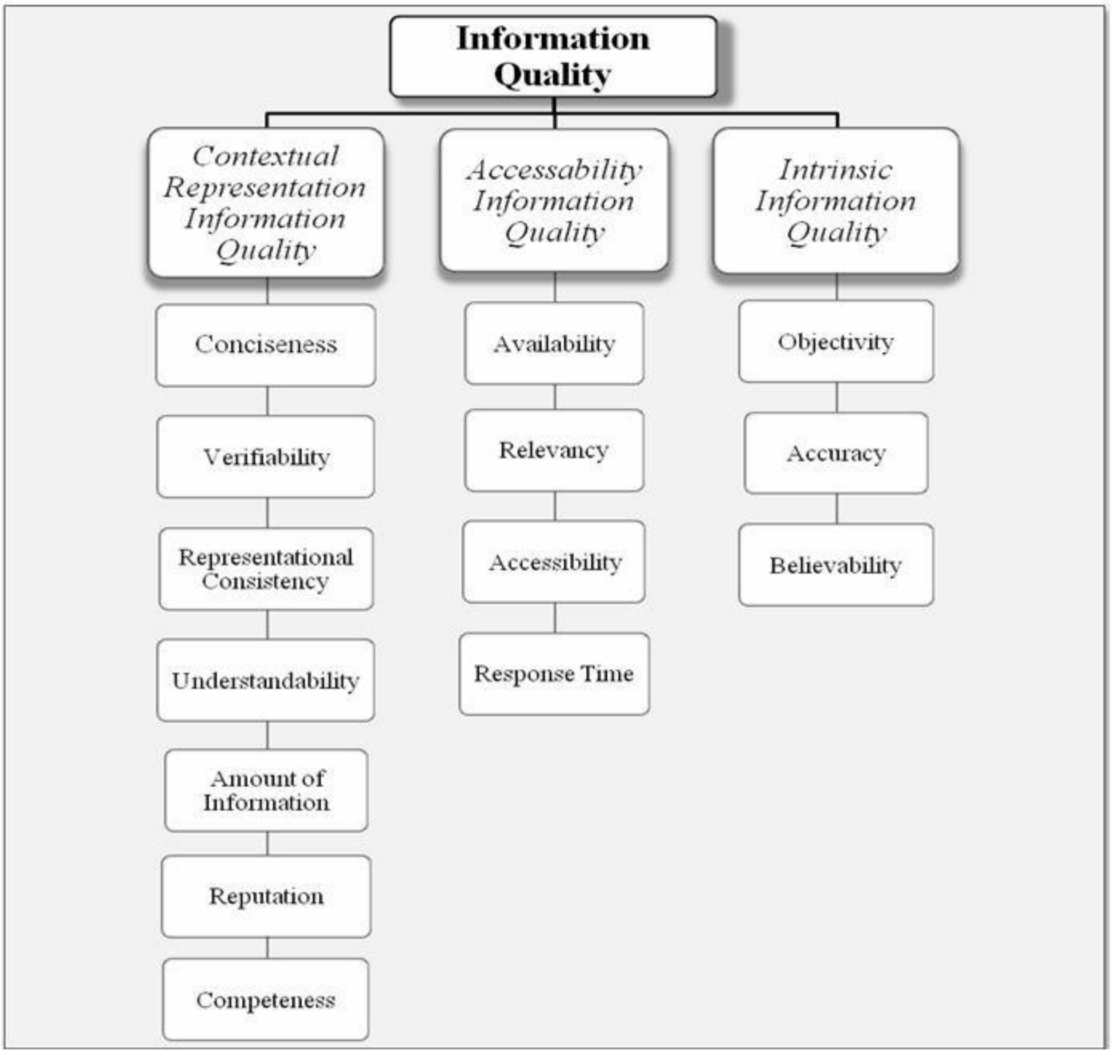


**Figure 4. The new proposed framework**

## 6. Revised framework

The revised framework for information quality in e-learning systems after calculating the relative importance for each dimension inside the three quality factors, and the relative importance for each factor in the overall quality are proposed in Figure 5. The final framework consists of 14 quality dimensions grouped in three quality factors: intrinsic, contextual representation and accessibility. The most important factor is Intrinsic information quality with relative importance score 41.157% of the overall quality while Contextual representation and Accessibility scored 33.851% and 24.992% respectively. Objectivity is the most important dimension in the Intrinsic factor. Reputation scored the highest relative importance within Contextual representation factor. Where Accessibility and Response time have almost the same relative importance within Accessibility factor with the scores 29.693% and 29.888% respectively.

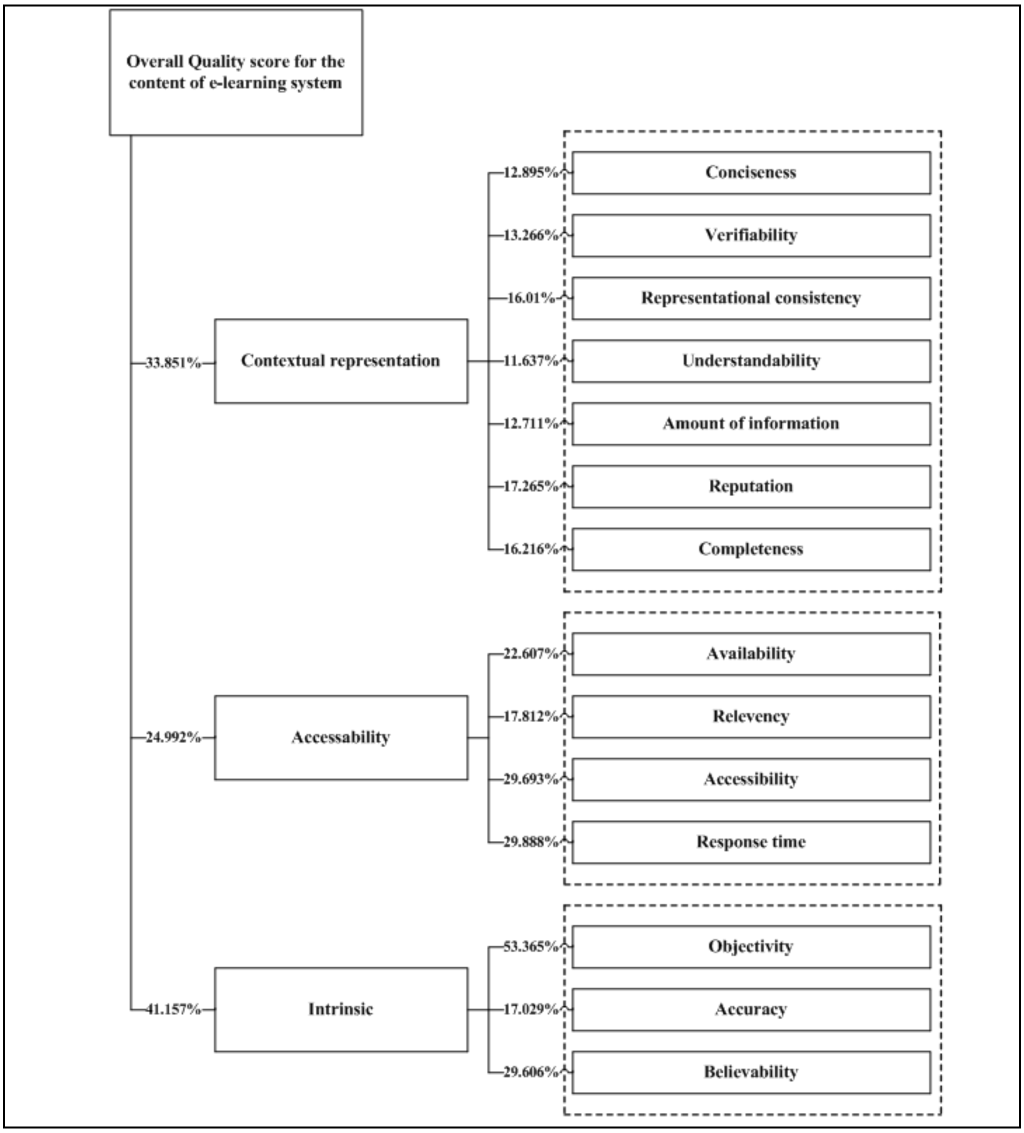


**Figure 5. The revised framework**

## 7. Conclusion and future work

Based on original questionnaire data and factor analysis, we proposed a new quality framework to measure the quality of the content provided by e-learning systems. Linear regression was used to calculate the relative importance for each dimension inside the three quality factors, and the relative importance for each factor in the overall quality. This framework could be used to provide a comprehensive indication of information quality in the context of e-learning systems. It could be useful to e-learning systems designers, providers and users as it provides a comprehensive indication of the quality of information in such systems.

As mentioned before, the framework proposed in this paper is a part of a larger research project. The next stage will be the development of a set of quality metrics and an experiment to compute these metrics in chosen e-learning systems. The value calculated for each metric will then be compared with the results from a user satisfaction survey. The research also will focus on taking advantages of software agent technologies in order to automate data collection and evaluation processes.

## Acknowledgements

The authors thank Professor Richard Wang, the director of the MIT Information Quality Program [MITIQ], for his kind permission to reproduce and use his data quality framework in this study.

Also would like to thank Professor Karl L. Wuensch from East Carolina University and Dr. Andy Field from Sussex University for their valuable help in applying the method to calculate the relative importance and linear regression.

## References

**1** Aladwani, A. M., & Palvia, P. C. (2002). Developing and validating an instrument for measuring user-perceived web quality. *Information & Management, 39*(6), 467-476.

**2** Alexander, J. E., & Tate, M. A. (1999). *Web wisdom: how to evaluate and create information quality on the web* (4th ed.). New Jersey *Lawrence Erlbaum Associates.*

**3** Besiki, S., Gasser, L., Twidale, M. B., & Smith, L. C. (2007). A framework for information quality assessment. *Journal of the American Society for Information Science and Technology, 58*(12), 1720-1733.

**4** Buyukozkan, G., Ruan, D., & Feyzioglu, O. (2007). Evaluating e-learning web site quality in a fuzzy environment. *International Journal of Intelligent Systems, 22*(5), 567-586.

**5** Chen, Y., Zhu, Q., & Wang, N. (1998). Query processing with quality control in the World Wide Web. *World Wide Web Journal, 1*(4), 241–255.

**6** Churchill, J., & Gilbert, A. (1979). A paradigm for developing better measures of marketing constructs *Journal of Marketing Research, 16*(1), 64-73.

**7** Close, R. C., Humphreys, R., & Ruttenbur, B. W. (2000). *e-learning & technology: technology & the internet are changing the way we learn*: Sun Trust Equitable Securities.

**8** Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd Revised ed.): *Lawrence Erlbaum Associates Inc*.

**9** Crisp, G. (2002). A model for the implementation and sustainability of a course management system in a research university. *the 19th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education*, Auckland, New Zealand:.

**10** Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 31*.

**11** Cronbach, L. J. (1971). Test validation. *R.L. Thorndike(Ed.),Educational measuremnt, American council on education*, Washington DC.

**12** Dedeke, A. (2000). A conceptual framework for developing quality measures for information systems. *Proceeding of 5th International Conference on information Quality*.

**13** Doherty, P. (1998). Learner Control in Asynchronous Learning Environments. *ALN Magazine, 2*.

**14** Ehlers, U.-D., Goertz, L., Hildebrandt, B., & Pawlowski, J. M. (2005). *Quality in e-learning Use and dissemination of quality approaches in European e-learning*. Luxembourg: the European Quality Observatory.

**15** Eppler, M., & Muenzenmayer, P. (2002). Measuring information Quality in the Web Context: A Survey of State-of-the-Art Instruments and an Application Methodology. *International Conference on Information Quality*.

**16** Field, A. (2000). *Discovering Statistics Using SPSS for Windows*: Sage Publications Ltd.

**17** Gerhard, J., & Mayr, P. (2002 ). Competing in the E-learning Environment--Strategies for Universities. *Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS'02)*.

**18** Gertz, M. (1996). Managing Data Quality and Integrity in Federated Databases. *2nd Working Conference on Integrity and Internal Control in Information Systems*.

**19** Hair, J. F., Tatham, R. L., Anderson , R. E., & Black, W. (1998). *Multivariate data analysis : with readings* (5th edetion ed.). London: *Pearson Education*.

**20** Hutcheson, G., & Sofroniou, N. (1999). *The Multivariate Social Scientist*. London: Sage Publications Ltd.

**21** Jarke, M., & Vassiliou, Y. (1997). Data Warehouse Quality: A Review of the DWQ Project. *2nd Intl. Conf. on Information Quality* Cambridge, Mass.

**22** Juran., J. (1974). *The Quality Control Handbook* (3rd ed.). New York,: *McGraw-Hill*.

**23** Kahn, K., Strong, D., & Wang, R. (2002). Information Quality Benchmarks: Product and Service performance. *Communications of the ACM, 45*(4), 184 – 193.

**24** Katerattanakul, P., & Siau, K. (1999). Measuring information quality of web sites: Development of an instrument. *20th international conference on Information Systems*, Charlotte, North Carolina, USA.

**25** Klein, B. D. (2002). When do users detect information quality problems on the World Wide Web? *American Conference in Information Systems*.

**26** Leung, H. K. N. (2001). Quality metrics for intranet applications. *Information & Management, 38*(3), 137 – 152.

**27** Liu, X. W., & Han, S. L. (2005). Ranking fuzzy numbers with preference weighting function expectations. *Computers & Mathematics with Applications, 49*(11-12), 1731-1753.

**28** Mecella, M., Scannapieco, M., Virgillito, A., Baldoni, R., Catarci, T., & Batini, C. (2002). Managing Data Quality in Cooperative Information Systems. *the Confederated International Conferences DOA, CoopIS and ODBASE*.

**29** Pallant, J. (2005). *SPSS Survival Manual*. United Kingdom: Open University Press.

**30** Paulsen, M. F. (2002). Online Education Systems: Discussion and Definition of Terms. *NKI Distance Education*.

**31** Peter, J. P. (1979). Reliability: a review of psychometric basics and recent marketing practices. *Journal of Marketing Research, 16*, 6-7.

**32** Porter, M. (1991). Towards a Dynamic Theory of Strategy. *Strategic Management Journal, 12*(1), 954-1117.

**33** Psaromiligkos, Y., & Retalis, S. (2003). Re-evaluating the Effectiveness of a Web-based Learning System: A Comparative Case Study. *Journal of educational multimedia and hypermedia, 12*, 5-20.

**34** Redman, T. (1996). *Data quality for the information age* (2nd ed.). London: *Artech House*.

**35** Robinson, P., Shaver, R., & Wrightsman, S. (1991). *Criteria for scale selection and evaluation. Measures of personality and social psychological attitudes*. San Diego, CA.

**36** Sekaran, U. (2000). *Research methods for business: A skill-building approach* (3 ed.). New York: *John Wiley and Sons*.

**37** Shanks, G., & Corbitt, B. (1999). Understanding data quality: Social and cultural aspects. *Proceedings of the 10th Australasian Conference on Information Systems;*.

**38** Spencer, D., & Hiltz, S. R. (2001). Studies of ALN: An Empirical Assessment. *Proceedings of the 34th Hawaii International Conference on System Sciences( HICSS-34).*

**39** Stracke, C. M. (2006). Quality Standards for Quality Development in e-Learning: Adoption, Implementation and Adaptation of ISO/IEC 19796-1. *Q.E.D. - The Quality Initiative E-Learning in Germany. The National Project for Quality in e-Learning.*

**40** Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5.

**41** Zeist, R. H. J., & Hendriks, P. R. H. (1996). Specifying software quality with the extended ISO model. *Software Quality Journal, 5*(4), 273-284.

**42** Zhu, X., & Gauch, S. (2000). Incorporating quality metrics in centralized/ distributed information retrieval on World Wide Web. *23 international ACM SIGIR conference on research and development in information retrieval*, Athens, Greece.

# Appendix I

## Correlation Matrix [a]

| | | Believability | Accuracy | Reputation | Consistency | Relevancy | Objectivity | Completeness | AmountOfInformation | Verifiability | Interpretability | Understandability | RepresentationalConsistency | Conciseness | Accessibility | ResponseTime | Availability |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Correlation | Believability | 1.000 | .445 | .246 | .188 | .300 | .412 | .302 | .248 | .345 | .356 | .204 | .311 | .235 | .166 | .190 | .188 |
| | Accuracy | .445 | 1.000 | .255 | .367 | .369 | .442 | .304 | .212 | .234 | .353 | .320 | .274 | .204 | .210 | .308 | .225 |
| | Reputation | .246 | .255 | 1.000 | .410 | .199 | .378 | .462 | .300 | .455 | .345 | .395 | .332 | .422 | .193 | .246 | .158 |
| | Consistency | .188 | .367 | .410 | 1.000 | .387 | .318 | .343 | .344 | .266 | .411 | .450 | .424 | .333 | .380 | .318 | .323 |
| | Relevancy | .300 | .369 | .199 | .387 | 1.000 | .288 | .356 | .376 | .299 | .251 | .252 | .248 | .193 | .367 | .282 | .342 |
| | Objectivity | .412 | .442 | .378 | .318 | .288 | 1.000 | .247 | .290 | .285 | .253 | .239 | .249 | .144 | .169 | .238 | .122 |
| | Completeness | .302 | .304 | .462 | .343 | .356 | .247 | 1.000 | .493 | .447 | .343 | .396 | .297 | .443 | .389 | .354 | .241 |
| | AmountOfInformation | .248 | .212 | .300 | .344 | .376 | .290 | .493 | 1.000 | .587 | .416 | .388 | .319 | .456 | .390 | .237 | .280 |
| | Verifiability | .345 | .234 | .455 | .266 | .299 | .285 | .447 | .587 | 1.000 | .508 | .492 | .366 | .496 | .413 | .354 | .350 |
| | Interpretability | .356 | .353 | .345 | .411 | .251 | .253 | .343 | .416 | .508 | 1.000 | .468 | .370 | .442 | .367 | .432 | .441 |
| | Understandability | .204 | .320 | .395 | .450 | .252 | .239 | .396 | .388 | .492 | .468 | 1.000 | .449 | .548 | .380 | .365 | .386 |
| | Representational Consistency | .311 | .274 | .332 | .424 | .248 | .249 | .297 | .319 | .366 | .370 | .449 | 1.000 | .557 | .293 | .318 | .254 |
| | Conciseness | .235 | .204 | .422 | .333 | .193 | .144 | .443 | .456 | .496 | .442 | .548 | .557 | 1.000 | .347 | .249 | .339 |
| | Accessibility | .166 | .210 | .193 | .380 | .367 | .169 | .389 | .390 | .413 | .367 | .380 | .293 | .347 | 1.000 | .348 | .435 |
| | ResponseTime | .190 | .308 | .246 | .318 | .282 | .238 | .354 | .237 | .354 | .432 | .365 | .318 | .249 | .348 | 1.000 | .435 |
| | Availability | .188 | .225 | .158 | .323 | .342 | .122 | .241 | .280 | .350 | .441 | .386 | .254 | .339 | .435 | .435 | 1.000 |

a. Determinant = .002

# Appendix II

## Anti-image Matrix [a]

| | | Believability | Accuracy | Reputation | Consistency | Relevancy | Objectivity | Completeness | AmountOfInformation | Verifiability | Interpretability | Understandability | RepresentationalConsistency | Conciseness | Accessibility | ResponseTime | Availability |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Anti-image Correlation | Believability | **.828[a]** | -.252 | .035 | .114 | -.105 | -.231 | -.121 | .064 | -.142 | -.170 | .100 | -.159 | .003 | .055 | .100 | -.027 |
| | Accuracy | -.252 | **.874[a]** | .038 | -.117 | -.159 | -.237 | -.076 | .077 | .071 | -.107 | -.114 | .005 | .015 | .035 | -.085 | .014 |
| | Reputation | .035 | .038 | **.834[a]** | -.238 | .041 | -.243 | -.270 | .152 | -.253 | -.038 | -.039 | .002 | -.158 | .133 | .018 | .074 |
| | Consistency | .114 | -.117 | -.238 | **.868[a]** | -.171 | -.072 | .001 | -.103 | .206 | -.148 | -.163 | -.202 | .067 | -.148 | .000 | -.055 |
| | Relevancy | -.105 | -.159 | .041 | -.171 | **.890[a]** | -.040 | -.109 | -.162 | -.025 | .109 | .047 | -.029 | .092 | -.121 | -.027 | -.169 |
| | Objectivity | -.231 | -.237 | -.243 | -.072 | -.040 | **.838[a]** | .086 | -.152 | -.011 | .049 | -.015 | -.035 | .129 | -.006 | -.076 | .061 |
| | Completeness | -.121 | -.076 | -.270 | .001 | -.109 | .086 | **.881[a]** | -.247 | .010 | .065 | -.044 | .089 | -.150 | -.152 | -.181 | .091 |
| | AmountOfInformation | .064 | .077 | .152 | -.103 | -.162 | -.152 | -.247 | **.861[a]** | -.357 | -.107 | .017 | .011 | -.149 | -.043 | .109 | .027 |
| | Verifiability | -.142 | .071 | -.253 | .206 | -.025 | -.011 | .010 | -.357 | **.876[a]** | -.160 | -.167 | -.012 | -.070 | -.141 | -.090 | -.039 |
| | Interpretability | -.170 | -.107 | -.038 | -.148 | .109 | .049 | .065 | -.107 | -.160 | **.921[a]** | -.077 | .015 | -.094 | -.029 | -.179 | -.168 |
| | Understandability | .100 | -.114 | -.039 | -.163 | .047 | -.015 | -.044 | .017 | -.167 | -.077 | **.934[a]** | -.091 | -.231 | -.042 | -.054 | -.096 |
| | RepresentationalConsistency | -.159 | .005 | .002 | -.202 | -.029 | -.035 | .089 | .011 | -.012 | .015 | -.091 | **.876[a]** | -.381 | -.018 | -.140 | .074 |
| | Conciseness | .003 | .015 | -.158 | .067 | .092 | .129 | -.150 | -.149 | -.070 | -.094 | -.231 | -.381 | **.865[a]** | -.036 | .122 | -.120 |
| | Accessibility | .055 | .035 | .133 | -.148 | -.121 | -.006 | -.152 | -.043 | -.141 | -.029 | -.042 | -.018 | -.036 | **.919[a]** | -.058 | -.194 |
| | ResponseTime | .100 | -.085 | .018 | .000 | -.027 | -.076 | -.181 | .109 | -.090 | -.179 | -.054 | -.140 | .122 | -.058 | **.881[a]** | -.238 |
| | Availability | -.027 | .014 | .074 | -.055 | -.169 | .061 | .091 | .027 | -.039 | -.168 | -.096 | .074 | -.120 | -.194 | -.238 | **.881[a]** |

a. Measures of Sampling Adequacy[MSA]

# Appendix III

**Coefficients for the first factor [a]**

| Model | | Correlations | | |
|---|---|---|---|---|
| | | Zero-order | Partial | Part |
| 1 | Believability | .388 | .000 | .000 |
| | Accuracy | .353 | .000 | .000 |
| | Objectivity | .368 | .000 | .000 |
| | Reputation | .672 | 1.000 | .162 |
| | Relevancy | .393 | .000 | .000 |
| | Completeness | .716 | 1.000 | .157 |
| | Amount Of Information | .708 | 1.000 | .139 |
| | Verifiability | .764 | 1.000 | .142 |
| | Understandability | .714 | 1.000 | .133 |
| | Representational Consistency | .658 | 1.000 | .156 |
| | Conciseness | .777 | 1.000 | .140 |
| | Accessibility | .480 | .000 | .000 |
| | Response Time | .422 | .000 | .000 |
| | Availability | .396 | .000 | .000 |

a. Dependent Variable: Factor#1

**Coefficients for the second factor [a]**

| Model | | Correlations | | |
|---|---|---|---|---|
| | | Zero-order | Partial | Part |
| 1 | Believability | .284 | 1.000 | .000 |
| | Accuracy | .376 | -1.000 | .000 |
| | Objectivity | .276 | -1.000 | .000 |
| | Reputation | .276 | 1.000 | .000 |
| | Relevancy | .647 | 1.000 | .237 |
| | Completeness | .463 | -1.000 | .000 |
| | Amount Of Information | .438 | .987 | .000 |
| | Verifiability | .494 | -1.000 | .000 |
| | Understandability | .486 | -.992 | .000 |
| | Representational Consistency | .386 | .973 | .000 |
| | Conciseness | .395 | 1.000 | .000 |
| | Accessibility | .760 | 1.000 | .306 |
| | Response Time | .732 | 1.000 | .307 |
| | Availability | .758 | 1.000 | .267 |

a. Dependent Variable: Factor#2

**Coefficients for the third factor [a]**

| Model | | Correlations | | |
|---|---|---|---|---|
| | | Zero-order | Partial | Part |
| 1 | Believability | .768 | 1.000 | .327 |
| | Accuracy | .733 | 1.000 | .248 |
| | Objectivity | .849 | 1.000 | .439 |
| | Reputation | .387 | .000 | .000 |
| | Relevancy | .394 | .000 | .000 |
| | Completeness | .351 | .000 | .000 |
| | Amount Of Information | .324 | .000 | .000 |
| | Verifiability | .370 | .000 | .000 |
| | Understandability | .316 | .000 | .000 |
| | Representational Consistency | .347 | .000 | .000 |
| | Conciseness | .238 | .000 | .000 |
| | Accessibility | .224 | .000 | .000 |
| | Response Time | .302 | .000 | .000 |
| | Availability | .212 | .000 | .000 |

a. Dependent Variable: Factor#3