1-1-2009

Limitations of Weighted Sum Measures for Information Quality

Markus Helfert
Dublin City University, markus.helfert@dcu.ie

Owen Foley
Galway Mayo Institute of Technology, Owen.Foley@gmit.ie

Mouzhi Ge
Dublin City University, mge@computing.dcu.ie

Cinzia Cappiello
Politecnico di Milano, cappiell@elet.polimi.it

Limitations of Weighted Sum Measures for Information Quality

Markus Helfert
Dublin City University, Ireland
Markus.Helfert@dcu.ie
Mouzhi Ge
Dublin City University, Ireland
mge@computing.dcu.ie

Owen Foley
Galway Mayo Institute of Technology, Ireland
Owen.Foley@gmit.ie
Cinzia Cappiello
Politecnico di Milano, Italy
cappiell@elet.polimi.it

ABSTRACT

In an age dominated by information, information quality (IQ) is one of the most important factors to consider for obtaining competitive advantages. The general approach to the study of IQ has relied heavily on management approaches, IQ frameworks and dimensions. There are many IQ measures proposed, however dimensions in most frameworks are analyzed and assessed independently. Approaches to aggregate values have been discussed, by which foremost research mostly suggests to estimate the overall quality of information by total all weighted dimension scores. In this paper, we review the suitability of this assessment approach. In our research we focus on IQ dependencies and trade-offs and we aim at demonstrating by means of an experiment that IQ dimensions are dependent. Based on our result of dependent IQ dimensions, we discuss implications for IQ improvement. Further research studies can build on our observations.

Keywords

Information quality, IQ dimensions, IQ assessment, IQ Measurement

INTRODUCTION

Information quality (IQ) has been often defined as a measure for 'fitness for use' of information (Wang and Strong, 1996). This discussion follows the general quality literature by viewing quality as the capability to 'meet or exceed users' requirements.' The literature provides numerous definitions and taxonomies of IQ dimensions analyzing the problem in different contexts. Also, literature provides us with numerous case studies, investigating IQ in practice. Common examples of IQ dimensions are accuracy, completeness, consistency, timeliness, interpretability, and availability. Over the last decade, many studies have confirmed that IQ is a multi-dimensional concept (Ballou and Pazer 1985; Redman 1996; Wand and Wang 1996; Wang and Strong 1996; Huang et al. 1999) and its evaluation should consider different aspects.

Much research in recent years has been focused on IQ assessment. Researchers have developed many frameworks, criteria lists and approaches for assessing and measuring IQ. The frameworks most widely used have been recently documented and adopted by the International Standards Organizations (ISO) (ISO 2008). However, despite the increasing interest in this topic, little is known about the effects and relationships between different IQ dimensions. Knight and Burn (2005) point out that despite the sizeable body of literature available relatively few researchers have tackled quantifying some of the conceptual definitions. However, clear definitions and insight about the relationship between IQ dimensions is essential for developing suitable measurement approaches. Many researchers have proposed measures for IQ dimensions, often underlying a weighted aggregate of single values for IQ dimensions (Wang and Strong, 1996). This, yet practical but simple aggregation does not provide an exact quality measure if dimensions affect each other. Assuming dependent IQ dimensions, a new approach for the overall quality evaluation should be proposed. In this article, we aim to review this problem by examining the relationship and dependencies that exist between selected IQ dimensions within current IQ frameworks. We relate our examination to the traditional weighted aggregate of single values.

In this paper, we introduce a general analysis about dependencies among quality dimensions. In order to verify our analysis we focus on an experiment showing the impact of variations in the accessibility dimension on the other quality dimensions.

The paper is structured as follows. In Section 2 we reflect our work with related research and outline limitations of current approaches. By the means of an experiment, Section 3 describes the experiment setting and results. Section 4 discusses the impact of dependent IQ dimensions in the context of IQ assessment and improvement. Section 5 concludes the article and presents indications for further research.

RELATED WORK

The rational for this work originate from the observation that IQ dimensions are not independent. Indeed, this general observation is supported by literature, which provides us with indications that IQ dimensions are dependent. Dependent IQ dimensions however impact the way of measuring IQ assessments, and thus would require a revision of the traditional weighted aggregate of single values as IQ measure. In order to reflect our analysis with related work we review foremost research relevant to IQ dimensions and IQ measures.

Dependency of IQ Dimensions

Many researchers have indicated various relations between IQ criteria, such as timeliness and availability. Ballou and Pazer (1995) propose a framework to investigate the tradeoffs between accuracy and timeliness in the context of decision making. Redman (1996) points out that timeliness has an impact on accuracy. Ballou and Pazer (2003) model the utility and tradeoffs between completeness and consistency. Olson (2003) implies the relationship between accuracy and completeness and states that consistency is a part of accuracy. Cappiello et al. (2004) analyze the time-related accuracy and time-related completeness in multi-channel information systems. Amicis et al. (2006) propose a data-driven approach to analyze the dependency between syntactic accuracy and timeliness as well as the dependency of completeness and timeliness. In Table 1 we combined a list of common IQ criteria and relations described in literature. As the list indicates, various trade-offs of IQ dimensions can be assumed. However, most researchers merely propose the relations but do not further investigate the strength or impact of the relation.

Item 1	Item 2	Source
Timeliness	Accuracy	Eppler (2001) adapted, Ballou and Tayi (1999), Ballou and Pazer (2003), Scannapieco and Batini (2006)
Timeliness	Believability	Eppler (2001) adapted
Timeliness	Consistent representation	Scannapieco and Batini (2006) adapted
Timeliness	Completeness	Scannapieco and Batini (2006)
Completeness	Accuracy	Ballou and Tayi (1999),Cappiello Francalanci and Pernici (2003), Fisher et al. (2006)
Completeness	Consistent representation	Ballou and Pazer (2003), Scannapieco and Batini (2006) adapted
Completeness	Conciseness	Eppler (2001) adapted, Fisher (2006) adapted
Accessibility	Security	Huang, Lee and Wang (1999), Eppler (2001), Fisher et al. (2006)
Accessibility	Accuracy	Missier et al. (2003)

Table 1. Selected relationships of IQ criteria

Related to specific framework, Table 2 summarizes IQ frameworks outlining the dimensions associated with each framework. The most prominent frameworks in the field of information systems and IQ research were selected. As an example, in Section 3 we focus on dependencies related to the accessibility dimension, which is pertain in most prominent frameworks (noted in column three of Table 2).

Framework	Dimensions / Quality Category	Accessibility
Wang and Strong (1996) (A Conceptual Framework for Information quality)	Believability, Accuracy, Objectivity, Reputation, Value-added, Relevancy, Timeliness, Completeness, Appropriate Amount of Data, Interpretability, Ease of understanding, Representational consistency, Concise Representation, Accessibility, Access Security.	Accessibility, Access Security.

Zeist and Hendricks (1996) (Extended ISO Model)	Functionality, Reliability, Efficiency, Usability, Maintainability, Portability	
Alexander and Tate (1999) (Applying a quality framework in a Web environment)	Authority, Accuracy, Objectivity, Currency, Orientation, Navigation.	
Katerattanakul et al (1999) (IQ of individual web sites)	Intrinsic, Contextual, Representational, Accessibility.	Navigational Tools Provided.
Shanks and Corbitt (1999) (Semiotic-based framework for IQ)	Well defined / formal syntax, comprehensive, unambiguous, meaningful, correct, timely, concise, easily accessed, reputable, understood, awareness of bias.	Easily Accessed.
Dedeke (2000) (Conceptual framework for measuring IS quality)	Ergonomic Quality, Accessibility Quality, Transactional Quality, Contextual Quality, Representational Quality	Technical access, System availability, technical security, data accessibility, data sharing, data convertibility

Table 2. Frameworks that consider the accessibility dimension

In order to assign a specific value to IQ a variety of IQ assessment methodologies have been proposed over the last decade. In the following we provide an overview of five typical methodologies (Redman, 1996; Huang et al., 1999; Lee et al. 2002; Pipino et al., 2002; Stvilia et al., 2006). We compare these selected methodologies by following criteria: definition of IQ dimensions, classification of IQ dimensions, model, tool, aggregation of IQ values, and case study. Definition of IQ dimensions is to identify that IQ dimensions are defined from which perspective. Classification of IQ dimensions is used to compare the classification of dimensions in each methodology. Model is to demonstrate the theoretical basis of the methodology. Tool is used to validate the implementation of the methodologies. By aggregation of IQ values, we describe how single IQ measurements are aggregated. Case study concentrates on empirical feasibility of these methodologies. Using the criteria above, we can obtain the characteristics of each methodology. If the methodology is only applied to one IQ community, it is considered as specific methodology. If the methodology can be applied to both IQ communities, it is a generic methodology. If the case study is provided in the literature, we regard the study as a practical study otherwise it is theoretical. We summarize the five methodologies and its characteristics in table 4.

	Redman (1996)	**Huang et al. (1999)**	**Lee et al. (2002)**	**Pipino et al. (2002)**	**Stvilia et al. (2006)**
Definition	12 IQ dimensions are defined from the database community	16 IQ dimensions are defined from management community	15 IQ dimensions are defined from both communities	16 IQ dimensions are defined from both communities	22 IQ dimensions are defined from both communities
Classification	Conceptual view, data value and representation	Classification of Wang and Strong (1996)	Classification of Kahn et al. (2002)	Without classifications	Classification of Wang and Strong (1996)
Model	A step by step procedure adapted from statistical process control	Adopt Deficiency model of Wand and Wang (1996)	Adopt PSP/IQ model of Kahn et al. (2002)	The model combines subjective and objective assessment	The model consists of activity types, IQ Problems, and IQ taxonomy
Tool	DCI system	IQ assessment survey	IQ assessment survey	IQ assessment software	IQ assessment survey
Aggregation of IQ values	Weighted Average	Average value of IQ dimensions	Weighted Average	Weighted Average	Potential impacts
Case Study	Telstra Co. Ltd.	Appliance Company		1, Global Consumer Goods, Inc., 2, Data Product Manufacturing , Inc.	1, Simple Dublin Core 2, English Wikipedia

Conclusion	Specific, practical	Specific, practical	Generic, Theoretical	Generic, practical	Generic, practical

Table 3. Comparison of IQ assessment methodologies

Many researchers have proposed ways to aggregate single measures of IQ dimensions, often underlying a weighted aggregate of single values for IQ dimensions (Wang and Strong, 1996). Although, recently some researchers have attempted to propose to identify IQ value curves and trade-offs by analyzing the potential impacts of IQ (e.g. Stvilia et al., 2006), many researcher propose to measure the overall impact of IQ as weighted aggregate. Also, A principle measure of the weighed sum of all the criteria (IQC_i) is illustrated in Equation 1, as

$$IQ = \sum_{i=1}^{N} \alpha_i IQC_i \quad \text{where} \quad \begin{array}{l} \forall \alpha_i : 0 \le \alpha_i \le 1 \\ \sum_{i=1}^{N} \alpha_i = 1 \end{array}$$

Equation 1. Aggregate measure of IQ

The proposed approach of totaling an IQ value by the weighted aggregate of single values for IQ dimensions underlay obviously certain assumptions, and thus has consequences on the form of relationships between IQ dimensions. Let us illustrate a simple example of the relation between timeliness and accessibility. The fundamental question here is "Is it better to have timely but less restricted information access, or to have higher access restrictions with less timely information?". Indeed one could argue that we can priorities both dimensions and assign a value to its importance. Let us assume, that accessibility is valued with α_1= 0.3 and timeliness is valued with α_2= 0.7 (For this example, we only consider two dimensions and all other dimensions are kept constant). With the weighted sum measure we could measure IQ as

$$IQ = 0.3 measure(accessability) + 0.7 measure(timelines)[+others]$$

Obviously, while we could represent trade-offs between single (independent) dimensions with this approach, the weighted sum is not suitable to consider dependencies among dimensions. As indicated with a simple illustration in Table 4, independent or depended dimensions would lead to different results for the same increase in accessibility (from initially 0.7 to 0.9 by some measures to improve accessibility). In the situation of independent dimensions, timeliness is unaffected. However, the situation changes fundamentally with dependent dimensions. The increase in accessibility affects also the timelines dimension (e.g. in our illustration timeliness increases to 0.7), although we only increased accessibility. Consequently, for such situations the weighted sum approach might be not an adequate aggregation approach. Indeed, for cost-benefit considerations we would need to consider the form of dependencies among dimensions (Cappiello/Helfert 2008), in order to represent the overall effect of IQ improvements correctly. Recently researchers have applied the partial least squaes analyses to IQ models, which are robust to many of the distributional assumptions of other modes (Bovee, 2004). Nonetheless, the common representation of IQ as weighted sum requires further investigations as it usually assumes independent IQ dimensions.

		Accessibility	Timeliness	Overall IQ according to weighted sum
Initial Situation		0.7	0.6	0.63
Independent dimensions	**Increase in Accessibility**	0.9	0.6	**0.69**
Depended dimensions		0.9	0.7*	**0.76**

* The increase results from an increase in accessibility, due to the dependency between accessibility and timeliness

Table 4. Comparison of dependent and independent dimensions using the weighted sum aggregation

ANALYSING DEPENDENCIES OF IQ DIMENSIONS

Section 2 summarized indications for dependencies of IQ dimensions. Motivated by this observation we aim to analyze some dependencies among IQ dimensions. In order to develop a suitable scenario, initial aim of our research is to ascertain the extent and dependency of the relationship between accessibility and other dimensions. This requires an examination of cause and effect. Galliers (1992) identifies the experimental method as the most suitable in this situation. This approach allows for an identification of the relationship between variables via an experiment design. Field and Hole

(2003) further believe that the goal of experimental research methods is to establish cause-and-effect relationships between variables. It is considered an empirical rather than an interpretive approach. It can be conducted by laboratory experiments, field experiments or a combination of both. Our research conducted a field experiments in attempt make general statements applicable to real life situations.

In addition to the variety of IQ frameworks, most frameworks provide their own definitions for accessibility. The definition of accessibility is framework dependent and some frameworks do not even consider it as a dimension of IQ. Loshin (2001) describes it as the degree of ease of access to information as well as the breadth of access. Wang and Strong (1996) consider that access security is also an important concept that must be taken into account when considering the dimension. Batini and Scannapieco (2006) describe accessibility in terms of the ability of the user to access the data from his / her own culture, physical status / functions and technologies available.

Research Model and Assumptions

In order to measure the impact of accessibility on dimensions in a framework manipulation of the accessibility dimension is required. The research aims to ascertain the impact of the accessibility dimension with respect to IQ dimensions. The independent variable is the level of accessibility with the dependent variables being the other dimensions associated with the particular IQ framework. In order to assess the impact of accessibility as a dimension of IQ the context of the IS, the tasks in hand, the users and the IS architecture need to be taken into account. Leung (2001) and Naumann and Rolker (2000) suggest that quality frameworks must take account of these factors.

The hypothesis of dependent IQ dimensions is motivated from the discussion of related research and our observations. The discussion led us to the investigation of the dependencies among IQ dimensions. The general dependency among IQ dimensions is illustrated by the research model in figure 1. In order to focus our research, initially our research has chosen four of the most common IQ dimensions in order to examine the impact of accessibility. The four dimensions chosen are free-of error, completeness, consistency and timeliness. As the experiment describes, we aim to test the hypothesis by varying the accessibility level. The hypotheses contend that accessibility levels have an impact on each of the individual dimensions to varying degrees. Our experiment demonstrates the extent of the impact.

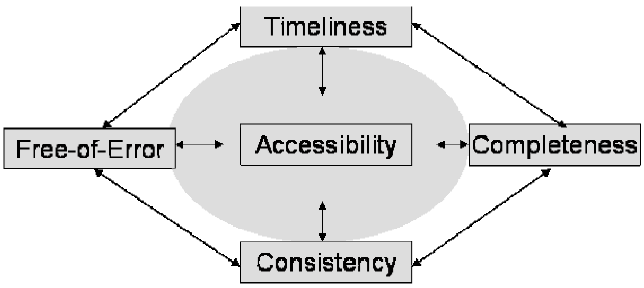

Figure 1. Research model

Experiment

Our hypotheses outlined above are validated by means of an experiment. Data can be collected in a number of ways in order to answer research questions. It can be gathered by direct observation or reported by the individual. Fisher et al. (2006) indicate that systematically collecting data to measure and analyze the variation of one or more processes forms the foundation of statistical process control. In the case of an experiment a variable is manipulated and the corresponding effect on the other variables is noted. Fisher et al. (2006) also point out that a statistical experiment is a planned activity where variables that have the potential to affect response variables are under the control of the researcher. For our research, we follow five distinct stages as outlined by Bernard (2000). The experiment examines four IQ dimensions across three architectures and two IS domains.

- *IQ Dimensions*: As IQ is a multidimensional concept the impact on individual dimensions is examined in the experiment. The research has selected four dimensions that are common across IQ frameworks free-of-error, completeness, consistency and timeliness.
- *Architectures*: Web, Client Server, Work Station
- *Domains:* The two IS domains are a library system and a student exam result system. The major areas of functionality of both systems are employed during the experiment. Three different access methods are used

namely workstation, client server and web. These are used on day to day operation of both systems. All users were also day to day operators of the systems. The pilot study focused on the Library IS.

The experiment manipulated the level of accessibility and measured the responses using the above ratings. There were 30 participants in the experiment, mainly bachelor and master level students with experiences in information systems. The accessibility level was randomly generated on a distribution between 0 and 100. The experiment was initially conducted with a subset of Wang and Strong's (1996) framework. The four dimensions chosen were Free-of-Error, Completeness, Consistency and Timeliness.

The following hypotheses are initially put forward based on the above initial dimensions above.

- H_1: Accessibility Level does impact the dimension Free-of-Error
- H_1: Accessibility Level does impact the dimension Completeness
- H_1: Accessibility Level does impact the dimension Consistency
- H_1: Accessibility Level does impact the dimension Timeliness

Varying Levels of Accessibility and Measuring its Effect

In order to create a response, we vary the level of accessibility in our experimental environment by adjusting the following components summarized in table 5. Lee et al. (2002) have proposed a number of metrics to measure dimension quality. These have been widely used to assess IQ. The experiment employs these metrics to measure the impact on the individual dimensions.

Accessibility Level	Implementation
Level of information accessibility	The percentage of queries that return the required information.
Level of system accessibility	The percentage of queries to which information system is available.

Table 5. Variations in Accessibility Levels

Free-of-Error	The dimension that represents whether data are correct.	Free-of-Error Rating = 1- $\left\langle \frac{N}{T} \right\rangle$ Where N = Number of data units in error and T = Total number of data units.	Count of the number of data units in error. Correct set of data as decided by custodian. In the case of the experiment Book Title and Student Name for Library and Student IS
Completeness	Schema, Column and Population	Completeness Rating = 1- $\left\langle \frac{C}{T} \right\rangle$ Where C = Number of incomplete items and T = Total number of items.	Degree to which entities and attributes are missing from the schema Book Title and its attributes along with Student and its attributes.
Consistency	Referential Integrity, Format	Consistency Rating = 1- $\left\langle \frac{C}{T} \right\rangle$ Where C = Number of instances violating specific consistency type and T = Total number of consistency checks performed.	Consistency between two related data elements. ISBN number and book. Student Number and Student Name were used.
Timeliness	The delay in change of real world state compared to the modification of the IS state. Redman (1996) defines as the difference between the times when the process is supposed to have created a value and when it actually has.	Timeliness Rating = R – I Where R = IS State Time I = Real World Time	Time difference between transaction commencement and change in IS state. Transactions such as add book, add student, alter book details and alter student details

Table 6. Response Measures

Experiment Results and Analysis

The results of the experiments are examined with respect to the correlation between the independent variable (accessibility level) and each (individual) dependent variable. Donnelly (2007) indicates that correlation measures both strength and direction of relationship between independent and dependent variable. The aim of this analysis is to ascertain what if any relationship exists between the variables.

Accessibility	Free-of-Error	Completeness	Consistency	Timeliness
83%	82%	78%	89%	**88%**
65%	84%	74%	86%	**62%**
45%	87%	56%	82%	**48%**

Table 7. Summary of Results

An examination of the initial hypotheses details the impact of accessibility on each of the individual dimensions. Table 7 summarizes our key results, which are illustrated as scatter plot by Figures 2 to Figure 5. An examination of the hypotheses reveals that relationships exist between accessibility and a number of the dimensions examined. The initial results indicate that there is a positive linear correlation between accessibility level and timeliness. As the accessibility levels to the information system increase, the timeliness dimension also improves. There is also a relationship between accessibility and completeness. At low levels of accessibility completeness is also low, however the relationship does not hold as levels of accessibility improve. At the higher level this relationship was again observed. This indicates fall off in the relationship and requires further examination. The scatter plot for both consistency and free-of-error did not display any relationship. The experiment indicates that accessibility level does not impact either of these dimensions.

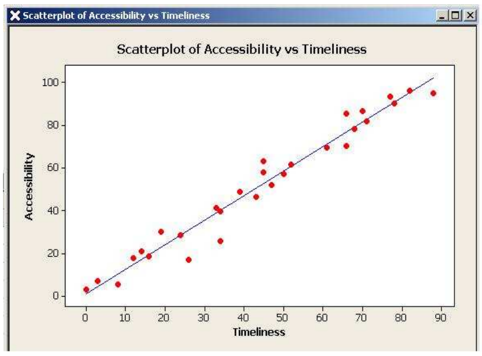

Figure 2. Accessibility vs. Timeliness

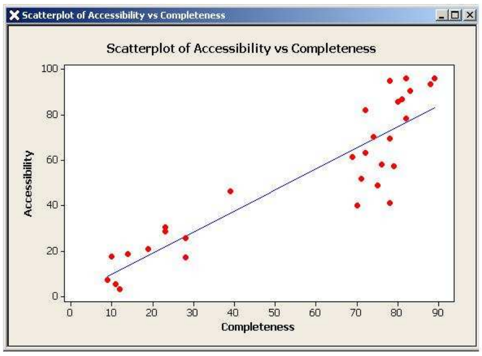

Figure 3. Accessibility vs. Completeness

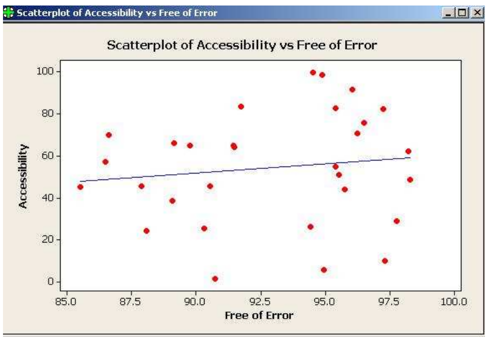

Figure 4. Accessibility vs. Free of Error

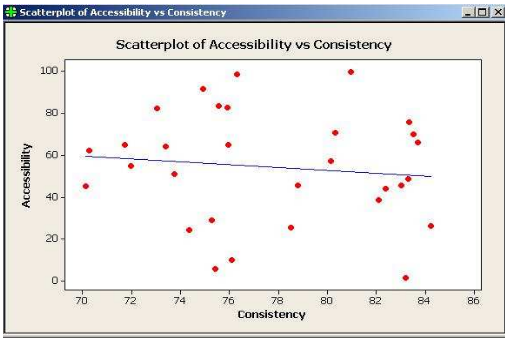

Figure 5. Accessibility vs. Consistency

IMPACTS OF DIMENSIONS DEPENDENCIES ON IQ ASSESSMENT AND IMPROVEMENT

The presence of dependencies among IQ dimensions affects the IQ assessment and improvement phases. Traditionally, as illustrated in Equation 1 dimensions are first assessed individually and then aggregated in order to obtain a concise IQ value. The aggregated value should define the quality level that characterizes information sources. As discussed above, the approach to use the average as aggregation functions is often not suitable among heterogeneous dimensions since dependencies introduces bias that negatively affect the reliability of the assessment procedure.

However, dependencies can be used in the improvement phase to improve its efficiency. In the improvement process, each improvement action can impact on a specific subset of quality dimensions. For example, data cleaning focuses on accuracy and consistency dimensions, data enrichment improves source completeness, source duplication is for data availability improvement and so on. In order to have a total data quality program, it is necessary to consider more than one action to increase the overall quality level and thus it is necessary to design the so called *improvement plan*. Dependencies among dimensions can be used as drivers for the selection of the improvement actions. Also we can consider dependencies among dimensions for the definition of the order with which actions should be executed. Due to the fact that if the quality dimension qd_1 depends on the quality dimension qd_2, improvements performed on the qd_2 would also increase the quality of qd_1. In our example timeliness is dependent on accessibility and it will benefit from each repair action that increases the accessibility level. The phenomenon is illustrated in Figure 2.

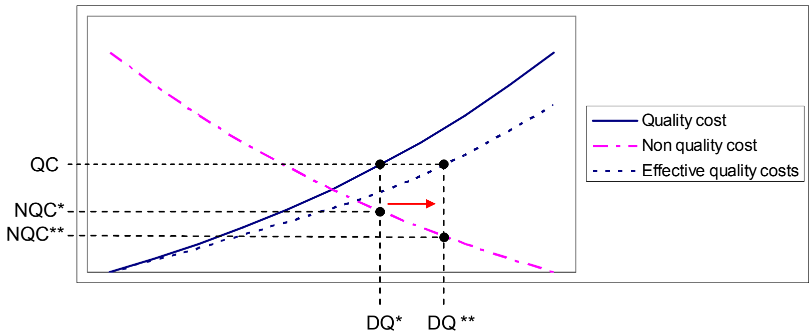

Figure 2. Dependencies benefits

Figure 2 illustrates cost curves associated with IQ. Non quality cost curve describes the trend of the costs associated with the process failures due to IQ problems. Non quality costs include for example, irrecoverable costs, analysis and correction costs, and re-execution costs. Quality costs are instead associated with all the improvement actions suitable for the specific context or system. Let us consider that the quality cost curve in Figure 2 regards costs related to an action able to improve accessibility dimension. Considering the single dimension, we can obtain a quality level DQ* having an amount of QC costs. If we consider dependencies, it is necessary to modify the quality costs curve since with the same amount of money, it is possible to obtain direct improvements on accessibility dimension and indirect improvements on timeliness dimension. In fact, the overall quality level is DQ** instead of DQ* and without difficulty we can verify that: *QC-NQC*<QC-NQC***. These considerations assist us in understanding the effect of dependencies among IQ dimensions. Our model can guide managers in defining the improvement plan and thus the schedule of the improvement actions in order to maximize the benefits. Furthermore, these considerations help us to understand the limitations of the weighted sum measure.

CONCLUSIONS AND FURTHER RESEARCH

Data quality research provided numerous methodologies to guide enterprise in the assessment, analysis, and improvement of data quality dimensions. Focusing on the critical issues related to the assessment phase, the literature does not provide an exhaustive set of metrics that organizations can apply. Several algorithms have been developed for a subset of dimensions, such as accuracy, completeness, consistency, and timeliness. The definition of an aggregate quality measure is still a much debated issue and existing contributions should be further analyzed and extended. In fact, as discussed in Section 2, the most common approach used to obtain a data quality index is to consider all the measures associated with the different quality dimensions and combine them by using a weighed sum. his approach has been criticized in this paper, since dependencies among data quality dimensions exist. Negative or positive dependencies affect the aggregate quality measure by introducing evident biases. Empirical research on one of the quality dimension (i.e., accessibility) has been conducted in order to verify these dependencies and to confirm our theories. However, dependencies do not introduce only criticisms in the assessment phase but they could be exploited to take advantages in the improvement phase. In fact, Section 4 shows that knowledge about positive or negative dependencies could drive the definition of an improvement plan. For example, by knowing that timeliness is positively influenced by accessibility, it is possible to schedule improvement activities focusing first on the actions that improve accessibility dimension in order to take advantages also on timeliness dimension. It could happen that benefits achieved by improving accessibility increases sufficiently the timeliness value and no further improvement actions are needed. In this way, through an appropriate improvement planning, organizations could maximize benefits and reduce costs. These first results will be further analyzed in the future work and more experiments are being planned in order to examine dependencies among a larger set of quality dimensions. Furthermore, future work will also focus on the definition of an algorithm to obtain an aggregate quality measure able to assess the organizations' data quality level. Finally, case studies will be considered for the validation of the proposed methods.

REFERENCES

1. Alexander, J. E., and Tate, M. A. (1999) Web Wisdom: How to Evaluate and Create Information Quality on the Web, Lawrence Erlbaum, Mahwah, NJ.
2. Amicis, F. D., Barone, D. and Batini C. (2006) An Analytical Framework to Analyze Dependencies among Data Quality Dimensions. In *Proceedings of the 11th International Conference on Information Quality*, MIT, USA.
3. Ballou D.P. and Pazer H.L.(1995) Designing Information Systems to Optimize the Accuracy-Timeliness Tradeoff, *Information Systems Research*, 6,1, 51-72.
4. Ballou, D. P. and Pazer, H. L. (2003) Modeling Completeness versus Consistency Tradeoffs in Information Decision Contexts. *IEEE Transactions on Knowledge and Data Engineering,* 15, 1, 240-243.
5. Ballou, D. P. and Tayi, G. K. (1999). Enhancing data quality in data warehouse environments. *Communications of the ACM*. 42, 1, 73-78.
6. Batini, C. and Scannapieco, M. (2006) *Data Quality Concepts, Methodologies and Techniques*, Springer, Heidelberg.
7. Bernard, H. R. (2000) *Social Research Methods, Qualitative and quantitative approaches* Thousand Oaks, Sage Publications, London, New Delhi.
8. Matthew Bovee (2004) Empirical Validation of the Structure of an Information Quality Model, *Proceedings of the Ninth International Conference on Information Quality (ICIQ-04)*, MIT, USA, 358-372.
9. Donnelly, R.A. (2007) The Complete Idiot's Guide to Statistics, Marie Butler-Knight.
10. Campbell, D. T. and Stanley, J. C. (1963) *Experimental and quasi-experimental designs for research* Houghton Mifflin, Boston.
11. Cappiello C., Francalanci C. and Pernici B.(2004) Time-Related Factors of Data Quality in Multichannel Information Systems, *Journal of Management Information Systems,* 20, 3, 71-91.
12. Cappiello, C. and Helfert, M. (2008) Analyzing Data Quality Trade-Offs in Data-Redundant Systems, in Alessandro D'Atri, Marco De Marco and Nunzio Casalino (Eds.) *Interdisciplinary Aspects of Information Systems Studies*, Physica-Verlag Heidelberg, 199-205.
13. Dedeke, A. (2000) A conceptual framework for developing quality measures for information systems. *Proceedings of 5th International Conference on Information Quality,* MIT, USA.
14. Eppler, M. and Muenzenmayer, P. (2002) Measuring Information in the Web Context: A survey of state of the art instruments and an application methodology in *5th International Conference on Information Quality*. MIT Boston, USA.
15. Field A. & Hole, G. (2003), How to Design and Report Experiments, Sage: London.
16. Fisher C.W. and Kingma B.R. (2001), Criticality of data quality as exemplified in two disasters, *Information & Management*, 39, 2, 109-116.
17. Fisher, C., Lauría, E., Chengalur-Smith, S. and Wang, R. (2006) Introduction to information quality.: M.I.T. Information Quality Programme, New York.
18. Galliers, R. (1992) Information Systems Research: Issues, Methods and Practical Guidelines. Blackwell Scientific.
19. Huang K. T., Lee Y. W., Wang R.Y.(1999), Quality Information and Knowledge Management, Prentice Hall.
20. ISO/IEC 25012 (2008) Software engineering -Software product Quality Requirements and Evaluation (SQuaRE) - Data quality model, International Organization for Standardization.
21. Jarvenpaa S. L. (1989) The Effect of Task Demands and Graphical Format on Information Processing Strategies, *Management Science*, 5,3, 285-303.
22. Kahn B.K. & Strong D.M. (1998) Product and Service Performance Model for Information Quality: An Update, in *Proceedings of the 1998 International Conference on Information Quality*, MIT, USA.
23. Kahn, B. & Strong, D. M. (2002) Information Quality Benchmarks: Product and Service. *Communications of the ACM*, 45, 4,184-192.
24. Katerattanakul, P. & Siau, K. (1999) Measuring information quality of web sites: Development of instrument, in *Proceedings of the 20th international conference on Information Systems*. Charlotte, North Carolina.
25. Knight, S. & Burn, J. (2005) Developing a Framework for Assessing Information Quality on the World Wide Web. *Informing Science*, 8, 159-172.
26. Lee Y., Strong D., Kahn B., and Wang R. Y. (2002), AIMQ: A Methodology for Information Quality Assessment, *Information & Management*, 40, 2, 133-146.
27. Leung, H.K.N. (2001) Quality Metrics for Intranet Applications. *Information & Management*, 38, 137-152.
28. Loshin, D. (2001) Enterprise Knowledge Management - The Data Quality Approach, Morgan Kaufmann.

29. Missier, P., Lalk, G., Verykios, V., Grillo, F., Lorusso, T. and Angeletti, P. (2003). Improving data quality in practice: a case study in the Italian public administration. *Distributed and Parallel Databases*. 13, 2, 125-160.
30. Naumann, F. & Rolker, C. (2000) Assessment methods for information quality criteria. *Proceedings of 5th International Conference on Information Quality,* MIT, USA.
31. Olson J. E. (2003), *Data Quality: The Accuracy Dimension*, Publisher: Morgan Kaufmann.
32. Pipino, L. Lee, Y.W. & Wang, R.Y. (2002) Data Quality Assessment. *Communications of the ACM*, 45, 4, 211-218.
33. Redman, T.(1996) Data Quality For The Information Age, Publisher: Artech House.
34. Savchenko, S. 2003. Automating Objective Data Quality Assessment, in *Proceedings of the 8th International Conference on Information Quality*, MIT, USA.
35. Shanks, G. & Corbitt, B. (1999) Understanding data quality: Social and cultural aspects. In *10th Australasian Conference on Information System.*
36. Stvilia, B., Gasser, L., Twidale M. B. and Smith L. C. (2007). A Framework for Information Quality Assessment, *Journal of the American Society for Information Science and Technology*. 58(12), 1720-1733
37. Wand Y., and Wang R. Y. (1996), Anchoring data quality dimensions in ontological foundations, *Communications of the ACM*, 39,11, pp.86-95.
38. Wang, R. Y. and Strong, D. M. (1996) Beyond accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information System*, 12,4, pp. 5-34.
39. Zeist, R.H.J. & Hendricks (1996) Specifying software quality with extended ISO model. *Software Quality Management*, 4, 145-160.
40. Zhu, X. & Gauch, C. (2000) Incorporating quality metrics in centralized/distributed information retrieval on the World Wide Web. *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*. Athens, Greece.