Factor Analysis of Critical Success Factors for Data Quality

Completed Research Paper

Hongjiang Xu
College of Business, Butler University
hxu@butler.edu

ABSTRACT

Data quality issues impact an organization's information system. Dirty data can damage every aspect of a business. In order to ensure data quality in information systems, it is important to understand the underlying factors that influence data quality. A research model was built based on the literature and previous case studies. In order to further develop and testing the research model for critical success factor for data quality in information systems, a large scaled survey was conducted and factor analysis were performed on the result of the survey. This paper provides a more scientific foundation for the research model of critical success factor for data quality. The study has theoretical and practical contributions to the field of data quality and information systems management.

Keywords

Data quality, Factor analysis, Critical success factors, Information systems.

INTRODUCTION

Data Quality

Companies lose billions of dollars annually due to poor data quality. Regardless of the organization size, data quality issues impact an organization's information system. With the proliferation of data warehouses, communication and information technologies have experienced an increase in the awareness of and need for high data quality (DQ) in organizations (Lee et al., 2002). Dirty data can damage every aspect of a business (D'Agostinoi, 2004.) Thus, DQ has been rated as a top concern to data consumers (Wang et al, 1998) and reported as one of the six categories commonly employed in management information systems research (Delone & McLean, 1992).

More and more electronically captured information requires processing, storage, and distribution through information systems (IS) (Siau et al., 2001). Advances in information technology (IT) have dramatically increased the ability and capability of processing accounting information. Real-world practice suggests that DQ problems are becoming increasingly prevalent (Huang, Lee & Wang, 1999; Redman, 1998; Wang & Wang, 1996).

Critical Success Factors for Data Quality

In order to ensure data quality in information systems, it is important to understand the underlying factors that influence data quality. Knowledge of the critical success factors (CSF) that constitute information systems having high data quality is desirable.

There have been many studies of critical success factors in quality management such as Total Quality Management (TQM) and Just-In-Time (JIT) (Saraph et al 1989; Porter & Parker 1993; Black & Porter 1996; Badri, Davis & Davis 1995). Some of the data quality literature has addressed the critical points and steps for DQ (Firth 1996; Segev 1996; Huang et al 1999; English 1999).

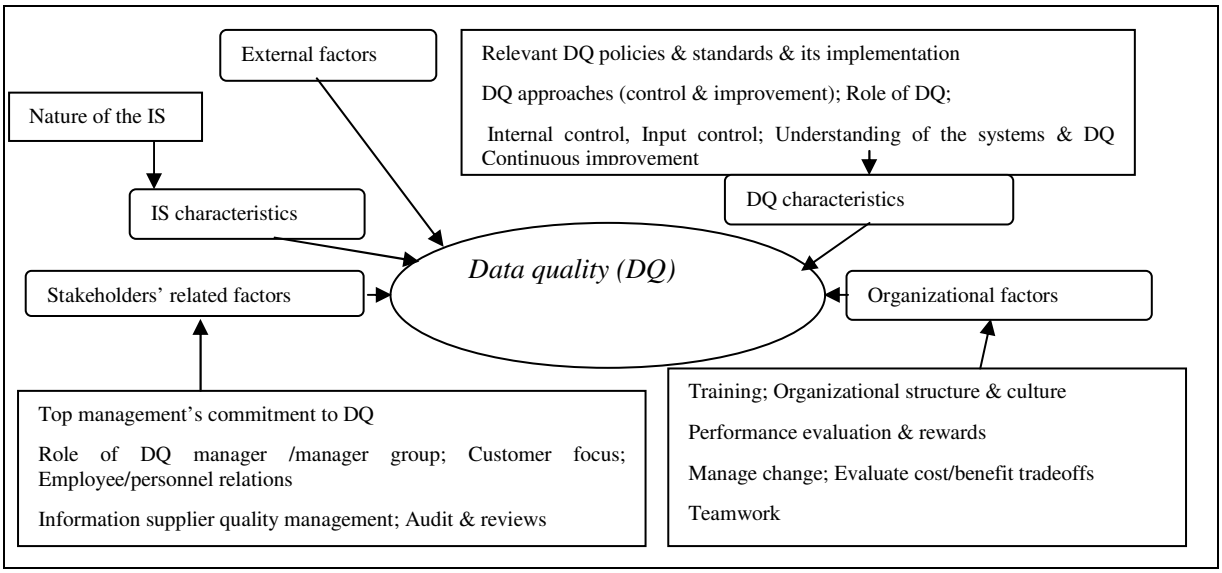
RESEARCH MODEL OF CRITICAL SUCCESS FACTORS FOR DATA QUALITY

The critical success factors model of information systems' data quality was developed, based upon the IS, data quality, quality management literature, and previous studies. Several categories of factors were identified that, according to the theoretical and empirical literature, have the potential to influence data quality in IS. These categories were IS characteristics, data quality characteristics, stakeholders' related factors, organizational factors, and external factors.

The relationship among factors and categories is shown in Figure 1 and forms the model for factors influencing data quality in information systems. There are seven factors listed under the category data quality characteristics; those factors are all related directly to the data quality itself. They are appropriate DQ policies and standards and its implementation, DQ approaches (control and improvement), Role of DQ, internal control, input control, understanding of the systems and DQ, and continuous improvement of DQ.

The stakeholders could come from both inside and outside the IS and the organization. Human related factors have always been the focus within social science and IT research. The category of stakeholders' related factors includes the human/people related factors' influence on DQ in information systems. They are top management's commitment to DQ, the role of DQ manager/manager group, customer focus, employee/personnel relations, information supplier quality management, and audits and reviews. At the organizational level, there are seven factors: training, organizational structure, organizational culture, performance evaluation and rewards, management of change, evaluation of cost/ benefit trade-offs, and teamwork (communication). External factors have been identified as factors outside the organization from the external environment, over which the organization has little or no control.

Figure 1: The Research Model for Factors Influencing Data Quality in Information Systems (modified from Xu et al., 2001, 2002)



However, this model was built based on the literature and previous case studies. In order to further develop and testing this model, a large scaled survey was conducted, and the result of the survey was used to perform factor analysis to identify factor groups from the 25 items from earlier research. A total of 1000 questionnaire was sent to professionals from different organizations. 180 questionnaires were completed and returned. SPSS and Lisrel software were used for data analysis and model testing.

FACTOR ANALYSIS

In order to group the 25 question items in the questionnaire into meaningful clusters, factor analysis is conducted via principal component analysis. Principal component analysis is used because it seeks to suitable rotation strategy; orthogonal varimax rotation is used because it minimizes the number of variables which have high loadings on any one given factor, resulting in easier identification of each variable with a single factor (Rennie, 1997). Apart from that, orthogonal rotation of question items also increases the generalizability of research findings (Rennie, 1997), which is deemed important for empirical research.

Only question items with factor loading of 0.40 and above were considered significant in interpreting the factors. None of the question items are removed because it does not correlate with any of factorial groups produced (Sadiq & Hoong, 2003).

Out of the 25 question items, four factors are produced. These factor groups addressed the criteria developed for the research study, as displayed in Figure 2. In selecting the number of factorial groups to be extracted, the KISer criterion is adopted.

KISer criterion, which is proposed in the 1960s, states that all components with eigenvalues under 1.0 are to be dropped. Thus, all the six factors which have values greater than 1 are extracted. These six factors accounted for 63.71% percent of the total variance.

Figure.2 Total Variance Explained

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	11.511	46.044	46.044
2	1.764	7.057	53.101
3	1.498	5.993	59.094
4	1.153	4.613	63.707
5	.917	3.668	67.374
6	.819	3.278	70.652
7	.658	2.631	73.283
8	.631	2.524	75.807
9	.592	2.366	78.173
10	.574	2.297	80.470
11	.524	2.095	82.565
12	.508	2.033	84.599
13	.472	1.889	86.488
14	.452	1.808	88.295
15	.401	1.606	89.901
16	.363	1.451	91.352
17	.333	1.334	92.686
18	.312	1.246	93.932
19	.286	1.145	95.077
20	.266	1.064	96.140
21	.263	1.051	97.191
22	.216	.864	98.055
23	.194	.777	98.833
24	.167	.667	99.499
25	.125	.501	100.000

Extraction Method: Principal Component Analysis.

These four factorial groups extracted are given the appropriate name in accordance with the criteria they represent from the proposed research model (see Figure 3). The name given for each factorial group and their associated criteria are:

(1) DQ Management Factors, representing the top/middle management in DQ, control and the change management.

(2) People & Assessment Factors, representing the effects of people including the employees, the suppliers, teamwork and so on.

(3) Organizational Factors, representing the organizational factors, such as the organizational culture and the organizational policy.

(4) Environmental & Personnel factors, representing the physical environment and personnel competency.

Figure.3 Cronbach alpha for factorial groups

Factor		Alpha scores	Revised alpha scores
1	DQ Management Factors	0.904	-
2	People & Assessment Factors	0.903	-
3	Organizational Factors	0.848	-
4	Environmental & Personnel factors	0.537	-

Cronbach's coefficient of reliability

The items in the factorial groups were also tested for reliability and validity. A reliability test was undertaken to ensure that the research findings has the ability to provide consistent results in repeated incidences (Sadiq & Hoong, 2003). To check the reliability aspect of the items and its factorial groups, internal consistency analysis using SPSS was performed. The items were grouped into its respective factorial group and coefficient alpha (Cronbach's) was calculated. The coefficients ranged between 0.537 for factor group "DQ Management" and 0.904 for factor group "Environment", which is below the value of 0.6 as suggested by Nunally (1967), suggesting weak internal consistency. Thus, the confirmatory factor analysis by Lisrel was conducted to further testing the model (See Figure 4). Also, out of the 25 question items used to measure DQ performances, no question items are eliminated because it showed no correlation with any factorial groups from the SPSS output. Therefore, no revised alpha scores are available for this Cronbach's coefficient of reliability.

Figure 4: Confirmatory factor analysis by Lisrel

Factor Group	**Performance Measure Items**	**Factor Loadings Grouping**			
		1	2	3	4
Data Quality Management Factors	Top management commitment	0.775			
	Middle management	0.755			
	Education & training	0.658			
	DQ vision	0.631			
	DQ control	0.581			
	Input control	0.707			
	User focus	0.674			
	Nature of IS	0.698			
	Change management	0.548			
People & Assessment Factors	Employee relation		0.569		
	Measurement report		0.677		
	Data supplier quality management		0.73		

		3	
	Continuous improvement	0.512	
	Teamwork	0.582	
	Evaluate cost / benefit tradeoffs	0.602	
	Understanding of the systems and DQ	0.568	
	Risk management	0.572	
	Audit & reviews	0.602	
	Internal Control	0.555	
Organizational Factors	Role of DQ Manager	0.751	
	Organizational structure	0.750	
	Policies & standards	0.662	
	Organizational culture	0.675	
Environmental & Personnel Factors	Personnel competency		0.790
	Physical environment		0.753

CONCLUSIONS

The research model developed based on existing literature and previous studies for critical success factor for data quality contained five factor category groups, and twenty five factor items. From the results of the factor analysis based on the large-scaled survey, the original research model of crucial success factor data quality in information systems should be modified with the four major factor category groups as: data quality management factors, people factors, organizational factors, and environmental factors; and the 25 factors items should be re-grouped under different factor groups. The factor analysis presented in this paper provided a more scientific foundation for the research model of critical success factor for data quality which is lacking from the existing literature. Thus the study has significant theoretical contribution to the field of data quality. In addition, it also provide guidance for practitioners in the data quality and information systems management fields of what are the critical success factors to ensure high quality data in their systems.

REFERENCES

1. Badri, M. A., Davis, Donald & Davis, Donna (1995), A study of measuring the critical factors of quality management, *International Journal of Quality and Reliability management*, vol. 12, no. 2, pp. 36-53.
2. Black, S. A. & Porter, L. J. (1996), Identification of critical factors of TQM, *Decision Sciences*, no. 27, pp. 1-21.
3. D'Agostino, D. (2004). Data Management: Getting Clean. CIO Insight, 1(42).
4. Delone, W. H., & McLean, E.R. (1992), Information system success: the quest for the dependent variable, *Information Systems Research*, *3*, 60-95.
5. English, L. P. (1999), Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits, John Wiley & Sons, Inc.
6. Firth, C. (1996), Data quality in practice: experience from the frontline, Paper presented to the Conference of Information Quality, 25-26 Oct.

7. Huang, Huan-Tsae, Lee, Y. W. & Wang, R. Y. (1999), *Quality information and knowledge*, Prentice Hall PTR.
8. Lee, W. Y, Strong, D. M., Beverly, K., & Wang, R. Y. (2002), AIMQ: A methodology for information quality assessment. *Information & Management*, *40*, 133-146.
9. Nunally, J.C. (1967), *Psychometric Theory*, McGraw-Hill, New York, NY.
10. Porter, L. J. & Parker, A. J. (1993), Total quality management- the critical success factors, *Total Quality Management*, no. 4, pp. 13-22.
11. Redman, T. C. (1998), The impact of poor data quality on the typical enterprise', *Communications of the ACM*, February vol. 41, no. 2.
12. Rennie, K.M. (1997), Exploratory and confirmatory rotation strategies in exploratory factor analysis, paper presented at the Annual Meeting of the South Educational Research Association.
13. Sadiq M & Hoong T. (2003), TQM practices and organizational performances of SMEs in Malaysia: Some empirical observations, *Benchmarking*. Vol.10. Issue. 1
14. Saraph, J.V., Benson, P. G. & Schroeder, R. G. (1989), An instrument for measuring the critical factors of quality management, *Decision Sciences*, vol. 20, no. 4, pp. 810-829.
15. Segev, A. (1996), On information quality and the WWW impact a position paper, conference of Information Quality, Oct 25-26.
16. Siau, K., Lim, E., & Shen, Z. (2001). Mobile commerce: Promises, challenges, and research agenda, *Journal of Database Management,* 12(3), 4-13
17. Wand, Y. & Wang, R. Y. (1996), Anchoring data quality dimensions in ontological foundations, *Communications of the ACM*, vol. 39, no. 11, pp. 86-95.
18. Wang, R. Y., Lee, Y. L., Pipino, L. & Strong, D. M. (1998), Manage your information as a product, *Sloan Management Review*, vol. 39, no. 4, pp. 95-105.
19. Xu, H., Koronios, A., & Brown, N. (2002), Managing Data Quality in Accounting Information Systems, In L. A. Joia (Ed.), *IT-Based Management: Challenges and Solutions*. Hershey, PA: Idea Group Publishing.
20. Xu, H., Koronios, A., & Brown, N., (2001), *A model for data quality in accounting information systems*. The invited session Data and Information Quality (DIQ), the 5th World Multiconference on Systemics, Cybernetics and Informatics (SCI'2001), Orlando, USA.