WHAT SKILLS MATTER IN DATA QUALITY?

(Research in-progress)

WooYoung Chung
Case Western Reserve University
wxc24@cwru.edu

Craig Fisher
Marist College
craig.fisher@marist.edu

Richard Wang
M.I.T. & U.C. Berkeley
rwang@mit.edu

Abstract: Research in data quality is a highly interdisciplinary field. Our community of data quality researchers and practitioners could benefit greatly from having a shared understanding of how diverse data quality research and skills complement one another, and to which area of data quality our attention should be directed. In this paper, we present a conceptual framework based on General Systems Theory, in order to facilitate collaborative discourse among data quality researchers and practitioners. In addition, we present an empirical measurement model based on General Systems Theory. We conducted an exploratory survey study at the 6th International Conference on Information Quality held at MIT in November 2001 (ICIQ-2001). Based on these survey data, we demonstrate how diverse data quality skills could be classified and contrasted. Our preliminary findings suggest that Adaptive Capabilities—the ability of identifying user requirements and measuring the user satisfaction and data quality—is perceived as most important. On the other hand, we also found that, among academics, executives, and managers, Interpretive Capabilities—the ability of identifying and articulating organizational implication of data quality—is considered as most important. In this paper, we discuss implications of these findings.

Key Words: Data Quality, Information Quality, TDQM, General Systems Theory, IS Curriculum

INTRODUCTION

Social construction of knowledge and skills has played a vital role in management science [2]. Indeed, no theory can simply describe the empirical realities as they are, nor any set of skills could be effectively used in all business situations. Hence, in order to identify which theories and skills are important for future research and curriculum development, management scientists have periodically taken stock of their research [6, 13, 19, 20, 43, 44]. They have also carefully surveyed and examined the collective sentiments of the community of management scientists and practitioners [33, 36, 42, 52]. In addition, a number of theoretical frameworks have been developed to describe how diverse management theories and skills fit together [3, 11, 22, 27, 28, 29, 41, 48]. All these studies have provided the opportunities for self-reflection within the management science community, facilitating social construction of cumulative knowledge in this field.

Research in data quality is no exception to this. In particular, data quality research is a complex interdisciplinary field spanning across diverse disciplines such as management, computer science, and psychology. We must draw upon knowledge and skills in these disparate disciplines, in order to conduct research in this area and design an effective curriculum for data quality professionals. Our community of data quality researchers and practitioners could benefit greatly from having a shared understanding of how diverse data quality research and skills complement one another, and to which area of data quality our attention should be directed.

For example, Wand and Wang (1996) describe ontological framework of data quality and provide an overview of past data quality research based on this framework [53]. In addition, Wang and Strong (1996) and Strong, Lee, and Wang (1997) examine how data "customers" define data quality [50, 54]. Studies like these facilitate the social construction of data quality research and they help us design effective curricula for data quality professionals.

Following this tradition, in this paper, we provide another theoretical perspective in which diverse data quality studies and skills could be organized. In particular, we adopt General Systems Theory as a conceptual framework [11, 12, 41]. General Systems Theory was originally developed to facilitate interdisciplinary academic discourse among highly disparate disciplines such as economics, physics, biology, and sociology. For this reason, this theory would be appropriate for conceptualizing the interdisciplinary research area such as the growing body of knowledge in data quality management.

In addition, in this paper, we present an empirical measurement model based on General Systems Theory. We conducted an exploratory survey study at the ICIQ-2001; we asked the conference participants to assess the relative importance of diverse data quality management skills. We present how these data quality skills could be classified according to General Systems Theory. Finally, in this theoretical perspective, we present how these participants evaluate relative importance of diverse data quality skills.

CONCEPTUAL FRAMEWORK

In General Systems Theory, Boulding (1956) suggests that diverse academic disciplines could be broadly classified into three levels, based on the nature and characteristics of the "system" that each discipline investigates: the mechanical system, the open system, and the human system levels [11]. This theoretical perspective has been widely accepted among management scientists [35]. For example, Morgan (1986) adopts this perspective and presents three metaphors for conceptualizing the organizational theories: Machine, Organism, and Brain [41]. Chaffee (1985) examines the strategy literature and classifies the studies of this area into three categories according to General Systems Theory: Linear, Adaptive, and Interpretive Strategies [12].

In this section, we describe how this framework could be adopted to describe the data quality research and skills. We refer to these three categories of data quality skills as *Technical Capabilities*, *Adaptive Capabilities*, and *Interpretive Capabilities*. In the following, we describe these three types of capabilities.

Technical Capabilities

Boulding (1956) points out that some disciplines focus on mechanical systems—the systems that have static structure and exhibit predetermined “clockwork” behaviors[11]. Typically, in these disciplines, mathematical modeling plays a central role in the epistemology. Mathematical models are used to examine complex dynamics within these mechanical systems. Boulding classifies disciplines such as physics, chemistry, and economics into this category.

In the Information Systems (IS) field, studies of diverse computational theories fit the description of the mechanical systems level (e.g., [10, 14, 16, 23, 25]). In particular, Chen (1976) and Codd (1970) provide an important foundation for modeling data [14, 16]. In addition, some studies on data quality extensively rely on mathematics and logical inference as their primary epistemological approach (e.g., [4, 5, 53]). These studies also fit into this category.

As for the practical skills, this category represents the skills of directly working with computer systems. For example, using relational algebra, data quality professionals may write SQL queries to identify incorrect or ambiguous data in databases. In addition, technicians may need the programming skills to write triggers and stored procedure to ensure the data integrity. These skills and knowledge are what we refer to as Technical Capabilities.

Adaptive Capabilities

The next level in General Systems Theory focuses on “open” systems. For example, biological living organisms interact with their environment, exchanging materials through ingestion, excretion, and diverse forms of metabolic exchange. Disciplines such as biology, physiology, and botany investigate how these open systems receive information from the outside, adapt to their environment, and effectively maintain the exchange process.

Data quality management could be considered as an open system [46, 54]. Data quality professionals need to interact with their environment—data users, managers, and other stakeholders. Their ability to effectively interact with these stakeholders is essential to data quality management. Many studies have examined diverse aspects of such interaction. For example, data quality professionals must be able to identify and define what these stakeholders want or need [22, 34, 46, 48, 54]. Voluminous studies of the user information satisfaction and data quality dimensions investigated how satisfaction of stakeholders could be conceptualized and measured [8, 9, 15, 24, 53]. Studies on technology acceptance also provide insight into how data quality professionals may encourage the use of information systems [1, 21, 35, 40, 49]. We refer to this ability and knowledge for effectively interacting with diverse constituents of data quality as Adaptive Capabilities.

Interpretive Capabilities

Unlike a biological organism, humans do not always simply react or adapt to their environment. As Boulding (1956) points out, humans are self-conscious self-reflective beings [11]. They interpret their situations and they assign symbolic social meaning to their actions. In addition, they are deeply embedded in dynamic social contexts. Their interpretations and symbolic actions influence their enactment of social structure in these social contexts and, at the same time, the social structure constraints their actions [26, 30]. These unique characteristics of human actions differentiate the human system from the open or mechanical systems.

Many studies have examined complex interaction between the use of information systems and human systems [6, 18, 45, 47]. For instance, Barley (1986) adopts the structuration theory to describe how a new technology could affect the patterns of interaction among people [6]. He demonstrates that a deployment of a new technology may serve as an impetus for restructuring social relationships in organization. In addition, Orlikowski and Yates (1994) also show that a new communication technology could bring about a new genre of communicative practice [45].

This ability of identifying and describing the complex interplay between technologies and organizational structure is what we refer to as Interpretive Capabilities. Data quality professionals should understand how data quality affects both formal and informal organizational structure—the way in which people interact and make decisions. Also, they should be able to articulate such implications of data quality to the top management and other stakeholders.

However, it is worth noting here that little systematic research has examined Interpretive Capabilities regarding data quality. So far, data quality research has focused primarily on improving the performance of individual decision makers. Organizational implications such as whether or not the availability of high quality data would bring about new ways of organizing have been neglected in the literature. For example, voluminous research supports the idea that managerial decision making is done in social contexts without “hard” data [17, 31, 32, 37, 38, 39, 49, 55]. Despite this finding, few research efforts have been made to investigate how such social structure, surrounding managerial decision-making, could be changed.

RESEARCH METHOD

At the beginning of the ICIQ-2001, we distributed copies of the four-page questionnaire to 110 conference participants. We received 61 usable questionnaires by the end of this two-day conference (the response rate is 55%). On average, our respondents were 40.2 years old and 27.9 percent of which were female. Our respondents had 15.7 years of work experience, on average; 10.5 years of this work experience was specifically related to IS.

Survey Instrument

Appendix 1 shows all data management topics listed in the questionnaire. This list of items is compiled based on our extensive literature review regarding data quality management topics, which includes, but not limited to, all items for Technical, Adaptive, and Interpretive Capabilities. The respondents are asked to rate these topics on a 7-point Likert-type scale, 1 indicating “Not at all important” and 7 indicating “Extremely important.”

Table 1 shows the descriptive statistics for all data management topics included in this study. This table also demonstrates how each questionnaire item is rated, in comparison with others, using T-Test and Tukey’s T-Test. The equivalent range columns indicate the items that are statistically the same based on these analyses. For instance, based on T-Test, item 1 is rated the same as item 2. Item 2 is rated the same as items 1 through 6 and 9.

No.	Description	Mean	Std.Dev	Equivalent Range for Each Topic: Based on T-Test	Equivalent Range for Each Topic: Based on Tukey	Subsets Based on Tukey			
1.	DQ measurement	6.33	0.96	1-2	1-10	A			
2.	DQ implications	6.10	1.08	1-6, 9	1-13	A	B		
3.	TQM	5.90	1.09	2-10	1-15	A	B	C	
4.	Data entry improvement	5.84	1.20	2-11	1-15	A	B	C	
5.	Org. policies	5.79	1.27	2-13	1-15	A	B	C	
6.	DB error detection	5.77	1.35	2-12	1-15	A	B	C	
7.	DQ dimensions	5.75	1.04	3-12	1-15	A	B	C	
8.	Change process	5.72	1.07	3-12	1-15	A	B	C	
9.	DQ cost/benefit	5.70	1.35	2-12	1-15	A	B	C	
10.	User requirements	5.67	1.14	3-14	1-15	A	B	C	
11.	Info. overload	5.49	1.21	4-15	2-15		B	C	
12.	DQ audit	5.46	1.18	5-15	2-15		B	C	
13.	Statistical techniques	5.30	1.46	5,10-15	2-18		B	C	D
14.	Data mining skills	5.23	1.50	10-15	3-18			C	D
15.	Data warehouse setup	5.18	1.43	11-15	3-18			C	D
16.	Analytic models	4.54	1.59	16-18	13-18				D
17.	Relational algebra	4.54	1.52	16-18	13-18				D
18.	Software tools	4.54	1.41	16-18	13-18				D

N = 61

TABLE 1. Descriptive Statistics and Comparisons among Topics

Measurement Model

The last column on the left in table 1 shows which items could be classified into subsets where all items are statistically the same. Based on Tukey, we identified four such subsets—these subsets are labeled A, B, C, and D in the last column. The interesting finding is that all items for Technical Capabilities ended up in subset D, the lowest group in our rankings list. On the other hand, the items that represent Adaptive and Interpretive Capabilities are included in subset A, the highest ranked group.

To test our measurement model, we conducted exploratory factor analyses using the principal component extraction method. Table 2 shows the final results of exploratory factor analysis. In this analysis, Varimax rotation of the final measurement model converged in 6 iterations.

Factor 1 represents Technical Capabilities in our framework. The reliability of this constructed, measured by Cronbach Alpha, is .798. Factor 2 is Adaptive Capabilities (Alpha = .644). Factor 3 represent Interpretive Capabilities (Alpha= .699). These constructs adequately demonstrate the convergent and discriminate validity of our measurement model.

	Factor 1	Factor 2	Factor 3
FACTOR 1 (ALPHA = .798)			
14. Data Mining Skills	.783		
16. Analytic Models	.765		
15. Data Warehouse Setup	.726		
17. Relational Algebra	.588	.508	
13. Statistical Techniques	.577	.442	
FACTOR 2 (ALPHA = .644)			
1. DQ Measurements		.813	
3. TQM		.675	.413
4. Data Entry Improvement		.612	
10. User Requirements		.553	
FACTOR 3 (ALPHA = .699)			
8. Change Process			.795
2. DQ Implications			.694
9. DQ Cost/Benefit			.644
6. DB Error Detection	.466		.561

Varimax rotation converged in 6 iterations.

TABLE 2. Factor Analysis

FINDINGS

Table 3 shows descriptive statistics and correlations among all variables used in this study. Using this table, one can also compare the relative importance of three types of capabilities in our framework. Adaptive Capabilities has received the highest importance rating (5.96 out of 7). Interpretive Capabilities is next, and the lowest is Technical Capabilities.

Table 4 shows the results of six regression analyses that focus on the difference in perception among researchers and practitioners. For all six equations, the central independent variable is Practitioner, a categorical variable that indicates whether or not the respondent is a practitioner (coded as 1 for practitioner and 0 for researcher). Two additional independent variables, age and gender, are added as control variables. For the first three equations, the dependent variables are Interpretive Capabilities, Adaptive Capabilities, and Technical Capabilities, respectively. In these three equations, none of the beta coefficients for the Practitioner variable is statistically significant.

In the remaining three equations in table 4, we use the difference in ratings between two skill variables as a dependent variable. The dependent variable for the first equation is the difference between Interpretive and Adaptive Capabilities. In the second equation, the difference between Interpretive and Technical Capabilities is used as the dependent variable. The difference between Adaptive and Technical Capabilities is used for the third equation.

In the last equation, beta coefficient for Practitioner is statistically significant ($\beta = .357$; $p < .01$). This finding suggests that practitioners value Adaptive Capabilities over Technical Capabilities far more than researchers—practitioners gave higher ratings to Adaptive Capabilities and lower ratings to Technical Capabilities.

	Mean	SD	Correlations 1	2	3	4	5
1. Interpretive Cap.	5.79	0.92					
2. Adaptive Capabilities	5.96	0.76	.288*				
3. Technical Capabilities	5.03	1.07	.462**	.269*			
4. Practitioner	0.65	0.48	.098	.205	-.258		
5. Age	40.11	8.53	.088	.326*	-.097	.060	
6. Female	0.24	0.43	.180	.162	.353**	-.039	-.079

* p < .05; ** p < .01

TABLE 3. Correlations and Descriptive Statistics for Variables used in This Study

	Dependent Variables Each Skills Sets			Dependent Variables Difference between Two Skill Sets		
	Interpretive Capabilities	Adaptive Capabilities	Technical Capabilities	Interpretive-Adaptive	Interpretive-Techical	Adaptive-Technical
Practitioner	-.097	.193	$-.241^{+}$	$-.233^{+}$	.162	.357**
Age	.108	.330*	-.056	-.148	.153	.273*
Female	.184	.195	.339*	.022	.186	-.190
Adjusted R^2	-.004	.129	.138	.027	.040	.219
F-Ratio	.917	3.624*	3.835*	1.489	1.734	5.958**

Numbers shown in this table are beta coefficients.
Total Degrees of freedom for each regression equation is 50.
$^{+}$ p < .10; * p < .05; ** p < .01

TABLE 4. Difference in Skills Ratings between Researchers and Practitioners

Indeed, table 5 clarifies how the perceptions of researcher and practitioner are different. In this table, we identified 6 major job titles in our sample and divide them into four groups: 1) professors; 2) executives and managers; 3) consultants; and 4) project managers and analysts. Respondents in these four groups rate the importance of the skills quite differently. Professors, executives, and managers (the first two sets of columns) rate Interpretive Capabilities the highest, followed by Adaptive and Technical Capabilities. On the other hand, Consultants, project managers, and analysts (the last two sets of columns in the table) rate Adaptive Capabilities the highest, then Interpretive and Technical Capabilities.

These findings provide some support for the idea that among practitioners, their job situations influence what skills they perceive as important. Executives and managers, who are primarily responsible for monitoring and deciding what should be done about data quality, see Interpretive Capabilities as the most important set of skills. On the other hand, consultants, project managers, and analysts are responsible for converting user requirements to technical specification, and reporting the status of data quality to the management. They perceive Adaptive Capabilities as the most important set of skills. Professors' ratings of these skills are congruent with executives and managers' rankings.

	Professors (N=10)		Executives Managers (N=11)		Consultants (N=10)		Project Managers Analysts (N=15)	
	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
Interpretive Capabilities	6.08	0.69	5.95	0.66	5.38	1.32	5.97	0.82
Adaptive Capabilities	5.68	0.83	5.64	1.10	6.30	0.60	6.17	0.60
Technical Capabilities	4.66	1.13	4.38	1.10	4.84	1.48	5.21	0.93
1. DQ measurement	5.80	1.14	5.91	1.51	6.50	0.71	6.80	0.56
2. DQ implications	6.50	0.97	5.73	1.10	6.20	1.48	6.27	0.70
3. TQM	5.80	1.03	5.55	1.69	6.00	1.05	6.20	0.86
4.Data entry improvement	5.70	1.34	5.64	1.21	6.20	1.23	6.00	1.07
5. Org. policies	6.10	0.99	5.64	1.29	5.70	1.25	6.13	1.13
6. DB error detection	5.90	0.99	6.09	1.30	5.10	1.91	6.00	1.07
7. DQ dimensions	6.00	1.25	5.45	1.29	5.70	1.16	5.87	0.83
8. Change process	6.00	0.94	5.73	1.01	5.30	1.25	5.80	1.01
9. DQ cost/benefit	5.90	1.45	6.27	0.79	4.90	1.73	5.80	1.42
10. User requirements	5.40	1.43	5.45	1.21	6.50	0.71	5.67	1.11
11. Info. overload	5.50	0.97	5.45	1.29	5.70	1.77	5.80	0.86
12. DQ audit	5.30	1.16	5.36	1.57	5.90	0.99	5.40	1.12
13. Statistical techniques	4.80	1.48	5.09	1.81	5.00	1.83	5.67	1.18
14. Data mining skills	5.00	1.33	4.27	1.74	5.30	2.11	5.60	1.18
15. Data warehouse setup	5.40	1.35	5.00	1.73	5.00	1.76	5.07	1.44
16. Analytic models	4.40	1.51	3.55	1.75	4.20	1.87	4.87	1.19
17. Relational algebra	3.70	1.77	4.00	1.90	4.70	1.34	4.87	1.06
18. Software tools	4.60	1.78	4.55	1.44	4.00	1.56	4.47	1.41

TABLE 5. Skills Ratings by Job Title

DISCUSSION

In this paper we proposed General Systems Theory as a conceptual framework for classifying diverse data quality skills. We suggested that data quality research and skills could be broadly classified into three categories: Technical, Adaptive, and Interpretive Capabilities. Our preliminary findings support the empirical measurement model of this framework.

Another objective of our study was to assess the relative importance of diverse data quality skills. Our findings suggest that technical skills, such as relational algebra and statistical techniques, are perceived as the least important in improving the data quality. However, this finding should be interpreted with caution. We had a very small number of database administrators, programmers and other “technicians” in our sample. Our analysis results do not adequately represent the opinions of these “technical” people.

We also found that data quality professionals with different jobs, value diverse skills differently. In particular, executives and managers perceive that Interpretive Capabilities is most important in effectively managing the data quality. Consultants, project managers, and analysts, on the other hand, rate Adaptive Capabilities the highest. A possible explanation is that these subjective evaluations are influenced by what these respondents do at their job. For instance, executives and managers would primarily focus on interpretation and assessment of the implications of data quality for the organization. On the other hand, the role of consultants and analysts would center on collecting the data quality requirements from the users or measuring the data quality.

These findings suggest that, in order to design an effective curriculum, one must consider the short-term and long-term career aspirations of their student. For instance, students who plan to get a job as an analyst or project manager, may benefit most from a curriculum focusing on enhancing the adaptive skills such as identifying user requirements and measuring the user satisfaction and data quality. On the other hand, executives training programs should emphasize the interpretive capabilities such as ability to assess organizational implications of data quality.

CONCLUSION

Research in data quality is a highly interdisciplinary field. In paper, we presented a conceptual framework based on General Systems Theory, in order to facilitate collaborative discourse among data quality researchers and practitioners.

We found that practitioners value diverse skills differently depending on their job situations. This finding suggests that IS educators should design a data quality curriculum to fit the need of long-term and short-term career objectives of their student.

In addition, in our survey, both academic researchers and executives reported Interpretive Capabilities—the ability of identifying and articulating organizational implications of data quality—as most important in improving and maintaining the data quality. Indeed, little systematic research has been conducted to examine how data quality would affect the way in which people are organized and jobs are structured. We recommend that future research efforts should be directed to this area.

REFERENCES

[1] Adams, D. A., Nelson, R. R., and Todd, P. A. (1992), "Perceived Usefulness, Ease of Use, and Usage of Information Technology: A Replication," *MIS Quarterly*, 16(2), 227-247.

[2] Astley, W. G. (1985), "Administrative Science as Socially Constructed Truth," *ASQ*, 30, 497-513.

[3] Astley, W. G. and Van de Ven, A. H. (1983), "Central Perspectives and Debates in Organization Theory," *Administrative Science Quarterly*, 28, 245-273.

[4] Ballou, D. P. and Pazer, H. L. (1985), "Modeling Data and Process Quality in Multi-input, Multi-output Information Systems," *Management Science*, 31(2), 150-162.

[5] Ballou, D. P. and Pazer, H. L. (1995), "Designing Information Systems to Optimize the Accuracy-timeliness Tradeoff," *Information Systems Research*, 6(1), 51-72.

[6] Barley, S. R. (1986), "Technology as an Occasion for Structuring: Evidence from Observations of CT Scanners and the Social Order of Radiology Department," *Administrative Science Quarterly*, 31, 78-108.

[7] Barley, S. R. and Kunda, G. (1992), "Design and Devotion: Surges of Rational and Normative Ideologies of Control in Managerial Discourse," *Administrative Science Quarterly*, 37, 363-399.

[8] Baroudi, J. J., Olson, M. H., and Ives, B. (1986), "An Empirical Study of the Impact of User Involvement on System Usage and Information Satisfaction," *Communications of the ACM*, 29(3), 232-238.

[9] Baroudi, J. J. and Orlikowski, W. J. (1988), "A Short-Form Measure of User Information Satisfaction," *Journal of Management Information Systems*, 4(4), 26-44.

[10] Bohm, C. and Jacopini, G. (1966), "Flow Diagrams, Turing Machines and Languages with only Two Formation Rules," *Communications of the ACM*, 9(5), 366-371.

[11] Boulding, K. E. (1956), "General Systems Theory -- The Skeleton of Science," *Management Science*, 2(3), 197-208.
[12] Chaffee, E. E. (1985), "Three Models of Strategy," *Academy of Management Review*, 10(1), 89-98.
[13] Chandler, A. D. (1977). *The Visible Hand: The Managerial Revolution in American Business*. Cambridge, MA: Harvard University Press.
[14] Chen, P. P.-S. (1976), "The Entity-Relationship Model--Toward a Unified View of Data," *ACM Transactions on Database Systems*, 1(1), 9-36.
[15] Chin, W. W. and Newsted, P. R. (1995), "RESEARCH REPORT: The Importance of Specification in Causal Modeling: The Case of End-user Computing Satisfaction," *Information Systems Research*, 6(1), 73-81.
[16] Codd, E. F. (1970), "A Relational Model of Data for Large Shared Data Banks," *Communications of the ACM*, 13(6), 377-387.
[17] Cohen, M. D., March, J. G., and Olsen, J. P. (1972), "A Garbage Can Model of Organizational Choice," *Administrative Science Quarterly*, 17, 1-25.
[18] Constant, D., Sproull, L., and Kiesler, S. (1996), "The Kindness of Strangers: The Usefulness of Electronic Weak Ties for Technical Advice," *Organization Science*, 7(2), 119-135.
[19] Culnan, M. J. (1987), "Mapping the Intellectual Structure of MIS, 1980-1985: A Co-citation Analysis," *MIS Quarterly*, 11(3), 341-353.
[20] Culnan, M. J. and Swanson, B. E. (1986), "Research in Management Information Systems, 1980-1984: Points of Work and Reference," *MIS Quarterly*, 10(3), 289-302.
[21] Davis, F. D., Bagozzi, R. P., and Warshaw, P. R. (1989), "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," *Management Science*, 35(8), 982-1003.
[22] DeLone, W. H. and McLean, E. R. (1992), "Information Systems Success: The Quest for the Dependent Variable," *Information Systems Research*, 3(1), 60-95.
[23] Dijkstra, E. W. (1968), "Go To Statement Considered Harmful," *Communications of the ACM*, 11(3), 147-148.
[24] Doll, W. J., Raghunathan, T. S., Lim, J.-S., and Gupta, Y. P. (1995), "RESEARCH REPORT: A Confirmatory Factor Analysis of the User Information Satisfaction Instrument," *Information Systems Research*, 6(2), 177-188.
[25] Gane, C. and Sarson, T. (1979). *Structured Systems Analysis*. Englewood Cliffs, NJ: Prentice-Hall.
[26] Giddens, A. (1984). *The Constitution of Society*. Los Angeles, CA: University of California Press.
[27] Henderson, J. C., Rockart, J. F., and Sifonis, J. G. (1987), "Integrating Management Support Systems into Strategic Information Systems Planning," *Jounral of Management Information Systems*, 4(1), 5-24.
[28] Hirschheim, R. and Klein, H. K. (1989), "Four Paradigms of Information Systems Development," *Communications of the ACM*, 32(10), 1199-1218.
[29] Ives, B., Hamilton, S., and Davis, G. B. (1980), "A Framework for Research in Computer-Based Management Information Systems," *Management Science*, 26(9), 910-934.
[30] Joas, H. (1987). Symbolic Interactionism. In A. Giddens & J. Turner (Eds.), Social Theory Today . Stanford, CA: Stanford University Press.
[31] Kotter, J. P. (1982), "What Effective General Managers Really Do," *Harvard Business Review*, 60(6), 156-167.
[32] Kotter, J. P. (1990), "What Leaders Really Do," *Harvard Business Review*(May-June), 103-111.
[33] Lee, D. M. S., Trauth, E. M., and Farwell, D. (1995), "Critical Skills and Knowledge Requirements of IS Professionals: A Joint Academic/Industry Investigation," *MIS Quarterly*, 19(3), 313-340.
[34] Lee, Y. W. (1996), "Why Know-Why Knowledge Is Useful for Solving Information Quality Problems," in the Proceedings of *the Americas Conference on Information Systems*.
[35] Mathieson, K. (1991), "Predicting User Intentions: Comparing the Technology Acceptance Model with the Theory of Planned Behavior," *Information Systems Research*, 2(3), 173-191.
[36] Miner, J. B. (1984), "The Validity and Usefulness of Theories in an Emerging Organizational Science," *Academy of Management Review*, 9(2), 296-306.
[37] Mintzberg, H. (1990), "The Manager's Job: Folklore and Fact," *Harvard Business Review*(March-April), 163-176.
[38] Mintzberg, H. (1994), "The Fall and Rise of Strategic Planning," *Harvard Business Review*(January-February), 107-114.
[39] Mintzberg, H., Raisinghani, D., and Theoret, A. (1976), "The Structure of "Unstructured" Decision

Process," *Administrative Science Quarterly*, 21(2), 246-275.
[40] Moore, G. C. and Benbasat, I. (1991), "Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation," *Information Systems Research*, 2(3), 192-222.
[41] Morgan, G. (1986). *Images of Organization*. Newbury Park, CA: SAGE Publications, Inc.
[42] Nelson, R. R. (1991), "Educational Needs as Perceived by IS and End-User Personnel: A Survey of Knowledge and Skill Requirements," *MIS Quarterly*, 15(4), 503-525.
[43] Orlikowski, W. J. and Baroudi, J. J. (1991), "Studying Information Technology in Organizations: Research Approaches and Assumptions," *Information Systems Research*, 2(1), 1-28.
[44] Orlikowski, W. J. and Iacono, C. S. (2001), "Research Commentary: Desperately Seeking the "IT" in IT Research-A Call to Theorizing the IT Artifact," *Information Systems Research*, 12(2), 121-134.
[45] Orlikowski, W. J. and Yates, J. (1994), "Genre Repertoire: The Structuring of Communicative Practices," *Administrative Science Quarterly*, 541-574.
[46] Orr, K. (1998), "Data Quality and Systems Theory," *Communications of the ACM*, 41(2), 66-71.
[47] Pickering, J. M. and King, J. L. (1995), "Hardwiring Weak Ties: Interorganizational Computer-mediated Communication, Occupational Communities, and Organizational Change," *Organization Science*, 6(4), 479-486.
[48] Seddon, P. B. (1997), "A Respecification and Extension of the DeLone and McLean Model of IS Success," *Information Systems Research*, 8(3), 240-253.
[49] Sproull, L. S. (1984), "The Nature of Managerial Attention," *Advances in Information Processing in Organizations*, 1, 9-27.
[50] Strong, D. M., Lee, Y. W., and Wang, R. Y. (1997), "Data Quality in Context," *Communications of the ACM*, 40(5), 101-110.
[51] Taylor, S. and Todd, P. A. (1995), "Understanding Information Technology Usage: A Test of Competing Models," *Information Systems Research*, 6(2), 144-176.
[52] Trauth, E. M., Farwell, D. W., and Lee, D. (1993), "The IS Expectation Gap: Industry Versus Academic Preparation," *MIS Quarterly*, 17(3), 293-307.
[53] Wand, Y. and Wang, R. Y. (1996), "Anchoring Data Quality Dimensions in Ontological Foundations," *Communications of the ACM*, 39(11), 86-95.
[54] Wang, R. Y. and Strong, D. M. (1996), "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems*, 12(4), 5-34.
[55] Weisband, S. P. (1987). Instrumental and Symbolic Aspects of an Executive Information System. In S. Kiesler & L. Sproull (Eds.), Computing and Change on Campus (pp. 150-169). Cambridge, MA: Cambridge University Press.

APPENDIX 1. List of the Survey Questions

No	Short Description	Survey Questions
1.	DQ measurement	Skills and knowledge of measuring data quality (such as timeliness, accuracy, completeness, and consistency of data)
2.	DQ implications	Understanding pervasiveness of data quality problems and their potential impacts
3.	TQM	Ability to apply the Total Quality Management principles (such as continuous improvement) to data quality management
4.	Data entry improvement	Skills and ability to analyze and improve data entry process in order to maintain data quality
5.	Org. policies	Ability to establish and maintain organizational policies and rules for data quality management
6.	DB error detection	Ability to detect and correct errors in databases
7.	DQ dimensions	Ability to define and describe diverse dimension of data quality (such as relevancy, believability, accessibility, ease of understanding)
8.	Change process	Ability to manage the change process/transitions resulting from data quality management project
9.	DQ cost/benefit	Skills and ability to conduct cost/benefit analysis of data quality management
10.	User requirements	Ability to translate subjective user requirements for data quality into objective technical specification (such as use of Quality Function Deployment)
11.	Info. overload	Understanding the information overload that managers often face and ability to reduce information overload
12.	DQ audit	Ability to conduct data quality auditing (formal review, examination, and verification of data quality)
13.	Statistical techniques	Ability to apply statistical techniques to manage and control data quality
14.	Data mining skills	Data mining and knowledge discovery skills for analyzing data in a data warehouse
15.	Data warehouse setup	Ability to integrate multiple databases into an integrated data warehouse
16.	Analytic Models	Ability to apply diverse analytic models (such as regression model and multidimensional model) for data analysis
17.	Relational algebra	Skills and ability to apply relational algebra (such as SQL) to estimate the accuracy of data
18.	Software tools	Experience and ability to use diverse commercially available data quality software packages