「정보시스템연구」 제14권 제2호
한국정보시스템학회
2005년 12월, pp. 237~256

Data Warehousing, Contextual Data Quality, and Problem Solving Performance

Won-Jin Jung* · Yong-Tae Park**

<목 차>

Ⅰ. Introduction

A data warehouse (DW) is a repository of integrated data derived from various databases and legacy data sources (Watson et al., 2001). According to Watson et al. (2001), the size of a data warehouse is usually massive and the data warehouse typically stores a wide range of information that has been generated over long periods of time. Accordingly, because of a huge amount of data in a DW people may experience information overload. In addition, data in a data warehouse may vary in quality and relevance to decision tasks, which makes decision tasks difficult for decision makers. Organizations may find themselves bogged down by database management systems such as data warehouses containing a number of data errors. In addition, the task of fixing errors in database management systems is very difficult. Consequently, organizations where decision makers experience information quality problems may end up taking unnecessary risks by accepting impractical ideas and making errors in interpretation, or ignoring important ideas. Wang and Strong (1996) noted that poor data quality can have substantial negative social and economic impacts. According to a recent

* Dankook University, professor, jungw@dankook.ac.kr
** California State University, Fullerton, professor, Corresponding author, ypark@fullerton.edu

study, the economic and social damage from poor quality information costs billions of dollars (Strong et al., 1997).

In the knowledge management literature, data are viewed as simple facts that become information as they are combined into meaningful structures (Tuomi, 1999). In the IS literature, information quality is one of two major dimensions for evaluating the success of information systems (DeLone and McLean, 1992; DeLone and McLean, 2003) and decision quality is a function of information quality (Stephenson, 1985). While the effects of information quality and the importance of information have been reported in the IS literature, little has been learned about the impact of data quality (DQ)[1] on decision performance. Thus, the purpose of this study is to empirically examine the relationship between data quality (DQ) and decision performance.

According to Wang and Strong (1996), data quality must be considered within the context of the task at hand. Specifically, they define high quality contextual data as data that can add value because it is relevant, timely, complete, and appropriate in terms of amount. Strong et al. (1997) studied 42 data warehousing and data quality projects and found contextual DQ problems in practice. Their examination revealed three underlying causes for data consumers' complaints that available data does not support their tasks: missing data, inadequately defined or measured data, and data that could not be appropriately aggregated. Since it is not well understood whether these contextual DQ problems affect problem solving performance in decision making settings, it would be worth investigating the effects of contextual DQ. In addition, the three causes for data consumers' complaints could increase the level of task complexity, which in turn has a negative impact on decision performance. Therefore, this study, which is part of a larger study, empirically explores how contextual DQ and task complexity simultaneously affect decision performance.

Ⅱ. Theoretical background and research hypothesis

2.1 Data Warehousing

Watson et al. (2001) conducted a descriptive study by using a survey to learn about current practices in data warehousing. One hundred and six data warehousing managers and professionals provided data about their companies' data warehousing experiences. The findings provided interesting and valuable insights into current data warehousing practices,

1) The authors use the terms data quality and information quality interchangeably.

especially the benefits of data warehousing. Some of the findings are: data warehousing is primarily the domain of large firms, but even smaller ones have a data warehousing initiative. Data warehousing sponsorship usually comes from a business unit, and marketing and sales is the most common driver behind a data warehousing initiative. They also found that there is no dominant architecture for data warehousing. Most importantly for this study, they pointed out that the greatest realized benefit from data warehousing is "more and better data."

Emerging technologies integrated with data warehousing offer competitive advantages to organizations. New forms of data are being integrated with data warehouses to provide high quality contextual data to decision makers. Large blocks of text, pictures, graphics, and even sound are being added to traditional data (Watson et al., 2001). Newly implemented data warehouses that include more varied complex data types require specialized database features. During the past several years, querying a database for text, color, or sound has become a reality. Because these characteristics and features of a new generation data warehouse are in many ways very different than those of a typical database, they may be able to increase the quality of contextual data.

2.2 Data quality dimensions

Since users of information systems would trust information of high quality and thus, are more likely to rely on such information in their decision making (Low and Mohr, 2001), researchers have identified a variety of data quality dimensions in an effort to provide the users of information systems with high quality information. Traditionally, studies on data quality issues have focused on the content of information (Strong et al, 1997a; Wang and Strong, 1996) and emphasized product and service dimensions of information quality (Kahn et al., 2002). For example, Raghunathan (1999) used accuracy as a measure of information quality, while Clikeman (1999) identified such information quality dimensions as relevance, reliability, timeliness, and cost. In addition, other researchers identified timeliness, accuracy, format, clarity, sufficiency, currency, and contextual validity as data quality dimensions (Khalil and Elkordy, 2005 Lillrank, 2003; Seddon and Kiew, 1994).

However, the traditional approaches to data quality dimensions showed some limitations in evaluating data quality in the context of e-business where information is browsed through hypermedia systems (Kim et al., 2005). Thus, they argue that people must consider the traditional content-related data quality as well as other data quality dimensions such as *form* and *time* in order to minimize irrelevant information, cognitive overload, and disorientation problems in the context of e-business. More specifically, they argue that information quality constructs such as information access security, history maintenance, information delivery, and information currency construct must be considered in evaluating both the product and service aspect of information quality in the context of e-business.

2.3. Contextual data

Contextual information is useful for not only information retrieval (IR) functions (Brown and Jones, 2001; Brown and Jones, 2002), but also for browsing tasks (Dourish et al., 1993; Park and Kim, 2000). IR systems are concerned with the finding of information, often in the form of text documents (Brown and Jones, 2001). Brown and Jones (2002) asserted that the use of context information leads to improvements in precision and retrieval speed. Jul and Furnas (1997) also contended that contextual information plays an important role for effective information retrieval because each retrieval process takes place in a particular information environment and is tied to the specificity of the environment.

Dourish et al. (1993) studied two information systems, one paper based and the other electronic, managing similar information within the same organization. In addition to the fact that the availability of contextual information makes browsing much more productive (Dourish et al., 1993; Park and Kim, 2000), they also found that information retrieved from these systems is interpreted subjectively by individuals, and they further pointed to contextual information contributing to this interpretation: that is, they addressed the importance of contextual information, which causes the same information to be interpreted indifferent ways once retrieved. Since this interpretation is critical in decision making, it must be perceived to be correct and pertinent if information is to be of use to an individual. Thus, they emphasized that contextual information acts as a resource in the process of interpreting the information.

Wang and Strong (1996) developed a hierarchical framework that captures the aspects of data quality (DQ) that are important to data consumers: Intrinsic, Contextual, Representational, and Accessibility Data Quality. According to them, contextual DQ highlights the requirement that data quality must be considered within the context of the task at hand; that is, data must be *relevant, timely, complete*, and *appropriate* in terms of amount so as to add value. To examine contextual DQ problems in practice, Strong et al. (1997) studied 42 data quality projects from three data rich organizations: GoldenAir, an international airline; BetterCare, a hospital; and HyCare, a Health Maintenance Organization (HMO). They found three underlying causes for data consumers' complaints that available data does not support their tasks: missing data inadequately defined or measured data, and data that could not be appropriately aggregated.

Strong et al. (1997) addressed the issue of incomplete data due to operational problems and design. For example, at GoldenAir, mechanics sometimes failed to record part numbers on their work activity forms. Because transaction data was incomplete, the inventory database could not be updated, which in turn produced inaccurate records. However, they observed that this was tolerated because the primary job of mechanics is to service aircraft in a timely manner, not to fill out forms. They also found that because of systems design, the amount of data in BetterCare's database is small enough to be accessible, but complete enough to be relevant and add value to data consumer's tasks. As a result, they observed that data consumers occasionally complained about incomplete data.

As organizations increasingly adopt distributed data repositories (such as data warehouses), it seems clear that various kinds of valuable data can be dispersed across the information systems in an organization. Strong et al. (1997) found some contextual DQ problems caused by integrating data across distributed systems. For example, data consumers complained about inconsistent definitions and irrelevant data. They pointed out that these problems were caused by autonomous design decisions in each division. As a result, decision makers experienced information overload and difficulties in retrieving valuable information through these distributed systems.

Based on these views, it could be possible to infer that decision makers can benefit from high quality contextual data because it can increase the efficiency and effectiveness of browsing and retrieval processes, as well as information interpretation processes (Brown and Jones, 2001; Dourish et al., 1993; Jul and Furnas, 1997; Park and Kim, 2000). That is, if the system provides high quality contextual data (information), then problem solving performance may be improved due to the improved efficiency and effectiveness of browsing, retrieval, and interpretation processes for the information necessary to make decisions. Therefore, when a person is given high quality contextual data for the experimental problem solving tasks, a positive effect of contextual data quality on decision performance is expected. Based on the discussion above, the following hypotheses are proposed.

H1: Regardless of the levels of task complexity, subjects with high quality contextual data will require less time in solving the task than subjects with low quality contextual data.

H1a: Subjects with high quality contextual data will require less time in solving the task than subjects with low quality contextual data for a simple task.

H1b: Subjects with high quality contextual data will require less time in solving the task than subjects with low quality contextual data for a complex task.

H2: Regardless of the levels of task complexity, subjects with high quality contextual data will have a higher accuracy in problem solving than those with low quality contextual data.

H2a: Subjects with high quality contextual data will have a higher accuracy in problem solving than those with low quality contextual data for a simple task.

H2b: Subjects with high quality contextual data will have a higher accuracy in problem solving than those with low quality contextual data for a complex task.

2.4. Task complexity

Task complexity is defined as the degree of cognitive load or mental effort required to identify and/or solve a problem (Payne, 1976). According to Campbell's (1988) concept of task complexity, tasks that increase information load and information diversity are considered complex tasks. Wood (1986) suggests that complexity is a function of the number of acts that must be executed and the number of information cues that must be processed when performing a task. Thus, tasks are considered more complex as the number of acts and information cues increases. In an information retrieval context, task complexity increases as the number of potential solutions increases because decision makers must evaluate each potential solution if they want to get the most effective or accurate result (Newell and Simon, 1972). Rossano and Moak (1998) also suggest that mental workload increases as more data are evaluated and retained in working memory.

Multi criteria tasks have a set of alternatives and a set of criteria. As the number of alternatives and criteria increases, decision makers must process more information. As mentioned above, task complexity increases where there are more information cues to process (Wood, 1986). Previous research on task complexity has shown that as task complexity increases, task difficulty increases. As a result, decision makers take more time and produce less accurate outcomes (Campbell, 1988; Crossland et al., 1995; Speier and Morris, 2003). Therefore, multi criteria tasks are considered more complex and difficult than elementary tasks (Crossland et al., 1995; Jankowski, 1995). Based on the discussion above, the following hypotheses are proposed.

H3: Regardless of the levels of contextual data quality, subjects with a simple task will require less time in solving the task than subjects with a complex task.

H3a: Subjects with high quality contextual data for a simple task will require less time in solving the task than subjects with high quality contextual data for a complex task.

H3b: Subjects with low quality contextual data for a simple task will require less time in solving the task than subjects with low quality contextual data for a complex task.

H4: Regardless of the levels of contextual data quality, subjects with a simple task will make more accurate decisions than subjects with a complex task.

H4a: Subjects with high quality contextual data for a simple task will make more accurate decisions than subjects with high quality contextual data for a complex task.

H4b: Subjects with low quality contextual data for a simple task will make more accurate decisions than subjects with low quality contextual data for a complex task.

Ⅲ. Research methodology

To examine the effects of contextual data quality (DQ) and task complexity on decision performance, a laboratory experiment using a Web based data warehouse was conducted. Based on the two factors, contextual DQ quality (high vs. low) and task complexity (simple vs. complex), a 2 x 2 factorial design was implemented to test the hypotheses. Various attributes of data quality (e.g., aggregated data, missing (incomplete) data, and irrelevant data) were used to map to the data type. Each subject's decision performance was assessed based on predetermined measurement, and decision performance referred to solution time and the accuracy of problem solving. A Web based simple system to deliver the contextual data to the subjects was developed using the latest version of Web programming languages, Hyper Text Markup Language (HTML) and Practical Extraction and Report Language (PERL). The system developed for this experiment can be viewed as a surrogate of the data access tools that are being used in various functional areas in industry because the subjects accessed data through this system.

3.1. Dependent and independent variables

The dependent variable of this study is decision performance, which was operationalized as the accuracy of problem solving and solution time. Problem solving accuracy was measured by the number of correct answers to the problems given in the decision task.. That is, problem solving accuracy was measured by dividing the number of corrected answers by the number of total problems and expressing the result as a percent of the correct solution. For this study the accuracy of problem solving was defined as:

$$\frac{\text{Number of Correct Answers} * 100}{\text{Number of Total Problems}}$$

This study measured solution time as the total time in seconds the subjects required to select the best site from the candidates. That is, solution time was measured from the time when the subjects began working on the task until they recorded their solutions on their answer sheet and logged out of the system. Fisher et al. (2003) distinguished between time constraints and time pressure. According to them, a time constraint is a specific allotment for making a decision, while time pressure is a subjective reaction to the amount of time allotted. Researchers found some mixed results with respect to the effects of time pressure on decision making. Time pressure decreases decision accuracy (Zakay and Wooler, 1984) and can impair the performance of some decision makers more than others (Ahituv et al., 1998). On the other hand, Austin (2001) found that increasing the time pressure may increase quality in software development projects. According to Dukerich and Nichols (1991), time constraints may have more impact on decision making for novices than for sophisticated decision makers. Because time factors, pressure or constraints, affect decision making (Ahituv et al., 1998; Austin, 2001; Zakay and Wooler, 1984), subjects were not informed of any time expectation for this experiment.

The decision task created by Jarvenpaa (Jarvenpaa, 2003) was used for this laboratory experiment, with some minor adjustments. It asks subjects to select a site for the construction of a Chinese restaurant. While the complex task asked subjects to select a site from among five alternative sites in which to locate a Chinese restaurant, the simple task asked subjects to select a site from among three alternative sites. The complex task had five factors for each site, while the simple task had three factors for each site. The factors were very important in deciding where the restaurant should be located. The scores for the factors were predetermined.

The degree of task complexity was manipulated by the number of problems in the task. The task required simple arithmetic calculations based on the decision criteria (factors) and decision choices (alternative sites for the restaurant). Specifically, the simple task with 24 problems required subjects to sum scores over three years for each factor. After averaging the summed scores for each factor, subjects were asked to sum the average scores for each site. Finally, they were asked to select a site that overall performs the best from among three alternative sites. On the other hand, the complex task with 80 problems required subjects to average scores over three years for each factor. In addition, subjects were asked to assign a weight for each factor. After that, they were asked to evaluate the sites by pair wise comparison (always comparing two sites at a time) with the weighted scores and select the site that wins the last comparison by having the largest number of factors of higher weighted value. That is, subjects were requested to rank the sites according to the predefined decision rules and the weighted scores for each factor.

Attaining high quality contextual data is a research challenge (Madnick, 1995; Strong et al., 1997), because tasks and their contexts vary across time and data consumers (Wang and Strong, 1996). Strong et al. (1997) found three main causes in general for data consumers' complaints that available data does not support their tasks: missing (incomplete) data, inadequately defined or measured data, and data that could not be appropriately aggregated. According to their findings, it seems possible to infer that providing data consumers with relevant, complete, and aggregated data may add value to the tasks of data consumers and may be one of the ways to solve the contextual DQ problems. In addition, according to the framework of data quality (Wang and Strong, 1996), one of the contextual DQ attributes is an appropriate amount of data. Therefore, providing problem solvers with an appropriate amount of data relevant to the tasks of data consumers could be another way to solve the contextual DQ problems.

The subjects supported with an appropriate amount of relevant, complete, and aggregated data in the form of tables were considered as being assigned to the experimental treatment of high quality contextual data. On the other hand, the subjects considered as being assigned to the treatment of low quality contextual data were given a limited amount of contextual data in the form of tables. That is, no aggregated data was given to them. In addition, they used irrelevant and incomplete data. For example, a couple of numbers in the data set given to the subjects was missing. Therefore, the subjects had to go through extra steps to infer the information necessary to make decisions.

3.2 Subjects

Subjects participating in the experiment were undergraduate student enrolled in various courses at California State University, Fullerton and California State Polytechnic University, Pomona. Classes were chosen from each of the universities and professors were asked to allow time for the experiment. After the classes were identified, time and location for the experiment were scheduled at participants' convenience. An orientation session was held one hour prior to the experiment to inform students about the experiment, providing a description of the experiment. Participation was entirely voluntary and students were informed that they would have the right to withdraw from the experiment at any time without penalty.

Since research has shown that incentives affect decision making effort and accuracy (Creyer et al., 1990), a performance based reward system was used for this experiment. A monetary incentive of $10 was awarded to all subjects after the experiment was completed. In addition, cash prices were awarded to subjects making the most accurate decisions in each treatment. The odds of winning were no worse than 1/10. In the case of ties, the subjects making the decisions in the least amount of time were determined as winners.

3.3. Experiment procedures and pilot study

The subjects were assigned randomly to one of the four treatments. The experimental task and a set of data were given to them. The task for this study asked subjects to solve a decision problem. In order to help subjects understand the decision making rules for the task, an example to simulate the decision making rules was provided. After that, the subjects were provided with an answer sheet to record their solutions as they performed the task. Next, with the data set and the task, the subjects made decisions. Finally, this study observed the effects of the various treatments on decision performance.

To test the experimental procedure and task complexity manipulations, two rounds of a pilot test were conducted. A total of eight undergraduate students participated. The results of the pilot study indicated that there were ambiguities in the decision making rules for the complex task. For the complex task, the subjects were asked to rank site alternatives from most desirable to least desirable, based on the decision rules defined in the task and factors that were weighted. That is, the decision rules and the weighted scores of the factors formed the basis for the subject's decision. The results indicated that the subjects assigned to the complex task were more likely to misunderstand the decision rules although an example to simulate the decision rules was provided. The examinations of the results led to two adjustments. First, subjects were instructed with a revised example to help them understand the decision rules easily. Second, to reduce the degree of task complexity for the complex task, the numbers of factors and alternative sites were reduced from 7 and 6 to 5 and 5, respectively.

Ⅳ. Data analysis and research findings

A total of 40 undergraduate students from various academic programs at California State University, Fullerton, and California State Polytechnic University, Pomona, participated in the experiment. Of the participants, 65 percent were male, and 60 percent were younger than age 25. The average age of participants was 24.6 years. The number of years in college was 2.8 years. Two thirds of the participants were majoring in business administration. Problem solving accuracy and time were each analyzed with two way ANOVA. The tests were carried out at a 95% confidence level. The descriptive statistics for the dependent variables are summarized in Table 1.

Table 1. Descriptive statistics for problem solving performance

Measures	Treatment Conditions			
	Simple Task		Complex Task	
	High Cont. DQ	Low Cont. DQ	High Cont. DQ	Low Cont. DQ
Solution Accuracy: (a higher score implies greater accuracy)				
Mean	97.917	47.500	97.917	74.444
Std. Dev.	4.0493	25.1692	1.6536	16.2993
n	10	10	10	10
Solution Time: (minutes: seconds)				
Mean	0:07:59	0:12:45	0:22:22	0:33:18
Std. Dev.	0:02:40	0:02:09	0:06:59	0:03:01
n	10	10	10	10

The interaction effect on problem solving time between task complexity and contextual DQ was significant (p = .029, see Table 2), indicating these two variables jointly affect problem solving time. That means more time was needed for subjects given low contextual DQ for both simple and complex tasks than for those given high contextual DQ. The explanation for these results probably lies in the low quality contextual data used. The low quality contextual data used in this study includes irrelevant and incomplete data. In addition, no aggregated data was provided. As a result, subjects using the low quality contextual data needed additional time to deal with the low quality contextual data as well as the decision tasks.

In addition, the results of the two way ANOVA for time showed that the main effects of task complexity (p = .000) and contextual DQ (p = .000) were significant (see Table 2). Since the interaction effect on problem solving time was significant, two one way ANOVAs were performed for these variables. The one way ANOVAs for time confirmed the significant main effects of task complexity (F = 108.549, p = .000) and contextual DQ (F = 17.712, p = .000). The results indicated that the simple task was solved more quickly than the complex task. Therefore, H3 was supported (see Table 4). Also consistent with expectations, subjects using high contextual DQ took less time than subjects using low contextual DQ. Therefore, H1 was supported (see Table 4).

Table 2. ANOVA for problem solving time

Source	Type III Sum of Squares	Mean Square	F	Sig.
Corrected Model	24092401	8030800	88.791	.000
Intercept	132295964	132295964	1462.71	.000
COMP	18018663	18018663	199.221	.000
CONT	5730386	5730386	63.357	.000
COMP * CONT	343351	343351	3.796	.029
Error	6873879	90445		
Total	163262245			
Corrected Total	30966280			

The two way ANOVA for problem solving accuracy revealed no significant interaction effect between complexity and contextual DQ (p = .869, see Table 3). However, the ANOVA on problem solving accuracy found a significant main effect for contextual DQ (p = .000, see Table 3). Therefore, H2 was supported (see Table 4). Surprisingly, the results showed that there was no significant main effect of task complexity for problem solving accuracy (p = .438, see Table 3). Subjects completing the complex task had comparable problem solving accuracy to those completing the simple task. Thus, H4 was rejected (see Table 4). This insignificant main effect of task complexity on problem solving accuracy might lead to the insignificant interaction between task complexity and contextual DQ despite the significant contextual DQ effect on problem solving accuracy.

Table 3. ANOVA for problem solving accuracy

Source	Type III Sum of Squares	Mean Square	F	Sig.
Corrected Model	18256	6085.347	7.658	.000
Intercept	271833	271833.472	342.069	.000
COMP	483	483.472	.608	.438
CONT	17750	17750.868	22.337	.000
COMP * CONT	21	21.701	.027	.869
Error	60395	794.674		
Total	350484			
Corrected Total	78651			

However, it is interesting to note that the main effect of task complexity was significant for problem solving time. These confounding results suggest that because there was no time constraint, that is, there was no specific allotment of time for making a decision, subjects used as much time as they needed to complete the complex task while keeping problem solving accuracy as high as possible. Table 4 summarizes the results of testing the hypotheses of this study.

Table 4. Summary of hypotheses testing

Hypotheses	Statistics		Evaluation
H1	F = 17.712	P = .000	Supported
H1a	F = 19.222	P = .000	Supported
H1b	F = 20.693	P = .000	Supported
H2	F = 22.337	P = .000	Supported
H2a	F = 39.112	P = .000	Supported
H2b	F = 18.474	P = .001	Supported
H3	F = 108.549	P = .000	Supported
H3a	F = 36.037	P = .000	Supported
H3b	F = 311.0	P = .000	Supported
H4	F = .608	P = .438	Rejected
H4a	F = .923	P = .349	Rejected
H4b	F = 3.775	P = .069	Rejected

V. Discussion of the findings and implications

According to the theory of cognitive fit (Vessey, 1991), when there is a complete fit of representation and task type, each representation (e.g., tables or graphs) will lead to both quicker and more accurate problem solving. Vessey (1991) states: "Spatial representations therefore best support the solution of spatial tasks; similarly, symbolic representations best support the solution of symbolic task (p. 227)." The task used in this study can be viewed as a symbolic task since it is best accomplished using precise data values. As mentioned previously, contextual DQ was found to influence problem solving performance. This indicates that although there was a cognitive fit between representation (e.g., tables) and task type (e.g., symbolic task), using high contextual DQ to solve the task resulted in more accurate outcomes than using low contextual DQ. In other words, when subjects used low contextual DQ, low contextual DQ detrimentally affected problem solving performance by increasing cognitive complexity, despite a fit between representation and task type. Dourish et al. (1993) found that information retrieved from information systems is interpreted subjectively by individuals. This subjective interpretation can be a critical determinant for the decision performance. Therefore, the inferior decision performance of subjects with low contextual DQ could be attributed to their subjective interpretations of the incomplete, irrelevant, and limited amount of data given to them.

Time pressure decreases decision accuracy (Zakay and Wooler, 1984) and can impair the performance of some decision makers more than others (Ahituv et al., 1998). Subjects with low quality contextual DQ were not given aggregated data and thus they actually had less time in solving problems than those with high quality contextual DQ. This probably put more time pressure on them and detrimentally affected their decision performance.

Based on a review of knowledge management literature, this research assumed that data is a prerequisite for information and information can be created from its raw data. DeLone and McLean (1992, 2003) postulated that system quality, information quality, and service quality singularly and jointly affect both system use and user satisfaction that are direct antecedents of "net benefits." Thus, based on DeLone and McLean's model and the assumption mentioned above, this study predicted that improved contextual data quality would positively affect information quality, which affects both system use and user satisfaction, which in turn have an impact on net benefits.

The results of this study showed that the effect of contextual DQ influences problem solving efficiency and effectiveness. Thus, the results of this study partially support the IS success model (DeLone and McLean, 1992; DeLone and McLean, 2003) in suggesting that information quality has an impact on user performance that is part of net benefits. That means, the analyses of research findings show that data quality as an antecedent of information quality has an impact on user performance. However, what is lacking is a detailed model for describing how data quality is transformed into information quality, the strength of the relationship between data quality and information quality, and the strength of the relationship between information quality and user performance. One area for future research would be to develop a model examining the transformation of data quality into information quality. Figure 1 presents a model for extending the IS success model by recognizing and including the contextual aspect of DQ into the model.

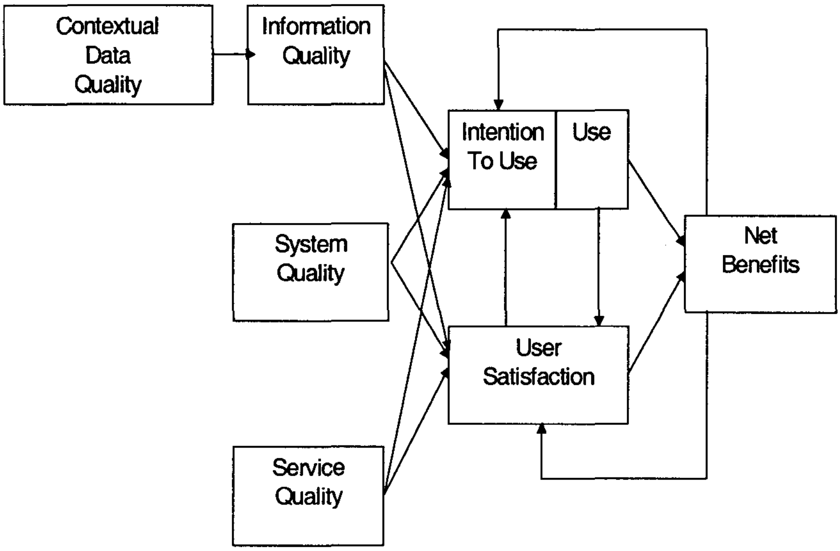

Figure 1. IS success model with contextual DQ

The observed main effect of contextual DQ on problem solving performance has practical implications for enhancing the efficiency and effectiveness of problem solving. In order to improve the decision performance of information system users, systems designers and managers should not only make data available to users, but also enable users to access high quality data. To accomplish this, it is recommended that systems designers and managers should support the task by providing users with high quality contextual data that are complete and relevant to the task. As organizations increasingly adopt distributed data repositories such as data warehouses, it seems clear that various kinds of valuable information can be dispersed across the information systems in an organization. Strong et al. (1997) found some DQ problems caused by integrating data across distributed systems. Thus, in order to enable users to access high quality data, systems designers, builders, and database administrators should ensure the integrity of data in data repositories such as data warehouses by implementing appropriate data cleansing and transformation rules and processes. Furthermore, since the contextual data required for a task may change over time, to improve decision performance of system users organizations must consider integrated database systems such as data warehouses that contain high quality contextual data, flexibly respond to users' request for data, and integrate data in ways that are required by different user groups in organizations (Strong et al., 1997; Wixom and Watson, 2001).

Ⅵ. Conclusion

This study explored the effects of contextual DQ and task complexity on decision performance. To examine the effects of contextual DQ and task complexity on decision performance, a laboratory experiment was conducted. The results demonstrated that the effects of contextual DQ on decision performance were significant. The findings suggest that decision makers can expect to improve their decision performance by enhancing contextual DQ..

This study adds to the decision performance literature by showing the effects of contextual data on decision performance. In addition, the findings of this study show that the cognitive fit theory and IS success model are valid in data quality context. In other words, the analysis of the findings show that using high contextual DQ as an antecedent of information quality, results in improved decision performance. Thus, these results confirm and extend the cognitive fit theory and IS success model. Finally, the findings of this study show that data quality is an important consideration in the design of database management systems.

A number of limitations should be considered in terms of the methods used when interpreting the findings. Since the study was conducted in laboratory settings with a small sample size of 40 undergraduate students and a specific task type, the outcomes of this study may not generalize to those of field experiments with a broader population. In addition, a single empirical study is not sufficient to validate the findings, further research should address these limitations and apply the findings of this study in specific contexts, population, and decision support technology as a whole.

References

Ahituv, N., Igbaria, M., and Stella, A., "The Effects of Time Pressure and Completeness of Information on Decision Making," *Journal of Management Information Systems*, Vol. 15, No. 2, 1998, pp. 153-172.

Austin, R.D., "The Effects of Time Pressure on Quality in Software Development: An Agency Model," *Information Systems Research*, Vol. 12, No. 2, 2001, pp. 195-207.

Brown, P.J., and Jones, G.J.F., "Context-aware Retrieval: Exploring a New Environment for Information Retrieval and Information Filtering," *Personal and Ubiquitous Computing*, Vol. 5, No. 4, 2001, pp. 253-263.

Brown, P.J., and Jones, G.J.F., "Exploiting Contextual Change in Context-Aware Retrieval," *Proceedings of the 2002 AGM Symposium on Applied Computing*, March 2002, pp. 650-656.

Campbell, D.J., "Task Complexity: A Review and Analysis," *Academy of Management Review*, Vol. 13, No. 1, 1988, pp. 40-52.

Creyer, E.H., Bettman, J.R., and Payne, J.W., "The Impact of Accuracy and Effort Feedback and Goals on Adaptive Decision Behavior," *Journal of Behavioral Decision Making*, Vol. 1, No. 1, 1990, pp. 1-16.

Crossland, M.D., Wayne, B.E., and Perkins, W.C., "Spatial Decision Support Systems: An Overview of Technology and a Test of Efficacy," *Decision Support Systems*,Vol. 14, No. 3, 1995, pp. 219-235.

DeLone, W.H., and McLean, E.R., "Information Systems Success: The Quest for The Dependent Variable," *Information Systems Research*, Vol. 3, No. 1, 1992, pp. 60-95.

DeLone, W.H., and McLean, E.R., "The DeLone and McLean Model of Information Systems Success: A Ten-Year Update," *Journal of Management Information Systems*, Vol. 19, No. 4, 2003, pp. 9-30.

Dourish, P., Bellotti, V., Mackay, W., and Ma C.Y., "Information and Context: Lessons from a Study of Two Shared Information Systems," *Proceedings of the Conference on Organizational Computing Systems*, December 1993, pp. 42-51.

Dukerich, J.M., and Nichols, M.L., "Causal Information Search in Managerial Decision Making," *Organizational Behavior and Human Decision Processes*, Vol. 50, No. 1, 1991, pp. 106-122.

Fisher, C.W., Chengalur-smith, I., and Ballou, D.P., "The Impact of Experience and Time on the Use of Data Quality Information in Decision Making," *Information Systems Research*, Vol. 14, No. 2, June 2003, pp. 170-188.

Jankowski, P., "Integrating Geographical Information Systems and Multi Criteria Decision-Making Methods," *International Journal of Geographical Information Systems,* Vol. 9, No. 3, 1995, pp. 251-273.

Jul, S. and Furnas, G.W., "Navigation in Electronic Worlds," *Nav 97 Report*, 1997.

Kahn, B., Strong, D., and Wang, R., "Information Quality Benchmarks: Product and Service Performance," *Communications of the ACM*, Vol. 45, No. 4, 2002, pp. 184-192.

Khalil, O.E.M. and Elkordy, M.M., "EIS Information: Use and Quality Determinants," *Information Resources Management Journal*, Vol. 18, No. 2, 2005, pp. 68-92.

Kim, Y.J., Koshore R., and Sanders L., "From DQ to EQ: Understanding Data Quality in the Context of E-business Systems," *Communications of the ACM*, Vol. 48, No. 10, 2005, pp. 75-81.

Lillrank, P., "The quality of information," *The International Journal of Quality & Reliability Management,"* Vol. 20, NO. 6/7, 2003, pp. 691-698.

Low, G.S. and Mohr, J., "Factors affecting the use of information in the evaluation of marketing communications productivity," *Academy of Marketing Science Journal,* Vol. 29, No. 1, 2001, pp. 70-88.

Lurie, N.H., "Decision Making in Information-Rich Environments: The Role of Information Structure," *Journal of Consumer Research,* Vol. 30, No. 4, 2004, pp. 473-486.

Madnick, S., "Integrating Information from Global Systems: Dealing with the "On-and Off-Ramps" of the Information Superhighway," *Journal of Organizational Computing,* Vol. 5, No. 2, 1995, pp. 69-82.

Newell, A., and Simon, H.A., "Human Problem Solving," Englewood, Cliffs, NJ: Prentice-Hall, 1972.

Park, J., and Kim, J., "Effects of Contextual Navigation Aids on Browsing Diverse Web Systems," *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, Vol. 1, No. 6, 2000, pp. 257-264.

Payne, J.W., "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis," *Organizational Behavior and Human Performance*, Vol. 16, No. 2, 1976, pp. 366-387.

Rossano, M.J., and Moak, J., "Spatial Representation from Computer Models: Cognitive Load, Orientation Specificity and the Acquisition of Survey Knowledge," *British Journal of Psychology*, Vol. 89, No. 3, 1998, pp. 481-497.

Seddon, P. and Kiew, M., "A partial test and development of the DeLone and McLean model of IS success," *Proceedings of the International Conference on Information Systems,"* 1994, pp. 99-109.

Speier, C. and Morris, M.G., "The influence of query interface design on decision-making performance," *MIS Quarterly*, Vol. 27, No. 3, 2003, pp. 397-423.

Stephenson, B.Y., "Management by Information," Information Strategy: *The Executive's Journal*, Vol. 1, No. 4, 1985 Summer, pp. 26-32.

Strong, D.M., Lee, Y.W., and Wang, R.Y., "Data quality in context," *Communications of the ACM*, Vol. 40, No. 5, 1997a, 103-110.

Strong, D.M., Lee, Y.W., and Wang, R.Y., "Decision Support for Exception Handling and Quality Control in Office Operations, *Decision Support Systems*, Vol. 8, No. 3, 1997b, pp. 217-227.

Tuomi, I., "Data Is More Than Knowledge: Implications of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational Memory," *Journal of Management Information Systems*, Vol. 16, No. 3, 1999, pp. 103-117.

Vessey, I., "Cognitive Fit: A Theory-Based Analysis of The Graphs Versus Tables Literature," *Decision Sciences*, Vol. 22, No. 2, 1991, pp. 219-241.

Wang, R.Y., and Strong, D.M., "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems*, Vol. 12, No. 4, 1996, pp. 5-34.

Watson, H.J., Annino, D.A., Wixom, B.H., Avery, K.L., Rutherford, M., "Current Practices in Data Warehousing," *Information Systems Management*, Winter 2001, pp. 47-55.

Wixom, B.H. and Watson, H.J., "An Empirical Investigation of the Factors Affecting Data Warehousing Success," *MIS Quarterly*, Vol. 25, No. 1, 2001, pp. 17-41.

Wood, R., "Task Complexity: Definition of the Construct," *Organizational Behavior and Human Decision Processes*, Vol. 37, No. 1, 1986, pp. 60-82.

Zakay, D., and Wooler, S., "Time Pressure, Training, and Decision Effectiveness," *Ergonomics*, Vol. 27, No. 3, 1984, pp. 273-284.

Jarvenpaa, S.L., "Additive-Difference Task: Ying-Yang Corporation Site Selection," Indiana University Kelley School of Business, (available online at http://kelley.iu.edu/bwheeler/ISWorld/index.cfm accessed Nov. 1, 2003).

<초 록>

Data Warehousing, Contextual Data Quality, and Problem Solving Performance

정원진 · 박용태*

데이터 웨어하우스는 기업의 통합된 데이터의 저장하는 곳이며, 대게는 상당한 규모를 가지고 있다. 또한, 데이터 웨어하우스는 일반적으로 다양한 종류의 데이터를 저장함으로 데이터 웨어하우스에 저장된 데이터는 의사결정 임무에 따라서는 그 질적, 적합성에 차이를 나타내고는 한다. 이러한 데이터 웨어하우스의 특성으로 인해서 때로는 데이터 웨어하우스의 데이터의 효용성이 기업의 의사결정을 지원하는데 있어 제한적일 수 있다. 정보 시스템의 문헌에는 데이터의 질이 의사결정 성과에 미치는 영향에 대해서 많이 알려져 있지 않다. 그래서, 본 연구는 contextual data(상황적 데이터)의 질과 업무의 복잡성이 의사결정 성과에 미치는 영향에 대해서 탐구해보고자 한다. Contextual data의 질과 업무의 복잡성이 의사결정의 성과에 미치는 영향을 조사하기 위하여 웹을 기반으로 하는 데이터 웨어하우스를 이용하는 실험을 실행했다. 연구의 결과는 contextual data의 질이 의사결정의 성과에 영향을 미친다는 것을 통계적으로 보여주었다. 이러한 연구결과는 의사결정자의 의사결정 성과를 향상시키기 위해서는 데이터 웨어하우스의 contextual data의 질을 향상시켜야한다는 것을 제시하고 있다.

색인어: 데이터의 질, 의사결정 성과, 데이터 웨어하우스

* 이 논문은 2005년 9월 20일 접수하여 1차 수정을 거쳐 2005년 10월 31일 게재 확정되었습니다.

* 교신 저자