Journal of the Association for Information Systems

Volume 4 | Issue 1 Article 6

8-1-2003

An Exploratory Investigation of System Success Factors in Data Warehousing

Bongsik Shin
San Diego State University, bshin@mail.sdsu.edu

Recommended Citation
Shin, Bongsik (2003) "An Exploratory Investigation of System Success Factors in Data Warehousing," *Journal of the Association for Information Systems*, 4(1), .
DOI: 10.17705/1jais.00033
Available at: https://aisel.aisnet.org/jais/vol4/iss1/6

Journal of the Association for Information Systems

An Exploratory Investigation of System Success Factors in Data Warehousing

Bongsik Shin
Department of Information and Decision Systems
San Diego State University
bshin@mail.sdsu.edu

ABSTRACT

Despite the increasing role of the data warehouse as a strategic information source for decision makers, academic research has been lacking, especially from an organizational perspective. An exploratory study was conducted to improve general understanding of data warehousing issues from the perspective of IS success. For this, the effect of variables pertaining to system quality, information quality, and service quality on user satisfaction for the data warehouse was studied. Additional characterization was made on data warehouse users, their organizational tasks, and data warehouse usage. Empirical data were gathered at a large enterprise from three different information sources: a survey, unstructured group interviews with end-users, and informal interviews with an IT manager who was in charge of the data warehouse. Data analysis showed that user satisfaction with the data warehouse was significantly affected by such system quality factors as data quality, data locatability, and system throughput. Interviews also supported the existence of system design and management issues that have to be addressed to optimize the utility of the data warehouse as an effective decision support environment. In the meantime, data analysis indicated that first-line (or lower) and middle managers were the main users of the system. Managers and knowledge workers were taking advantage of the system to perform complex tasks, to support decision making, and to seek information critical for enhanced productivity. The group interviews revealed additional benefits of the data warehouse and major roadblocks in its successful usage.

Keywords: Data warehouse, decision support system, IS success, critical success factors

I. INTRODUCTION

The data warehouse has emerged as a key platform for the integrated management of decision support data in organizations. Its potential for supporting business intelligence has been widely noted. This technology has become crucial in an environment where increasing competition, sophisticated and informed consumers, unpredictable market fluctuations, and changing regulatory environments are putting much pressure on business organizations. Organizations use the data warehouse for a variety of tasks such as planning, target marketing, decision making, data analysis, and customer services, and practitioners agree that it is changing the way business is done.

The data warehouse can improve business performance in various ways, including better-targeted products, improved customer relationship management, and greater operational efficiency [Cooper et al. 2000; Moore and Wells 2000]. It can also result in reengineering of business processes [Srivastava and Chen 1999]. For instance, automated and integrated information delivered from the data warehouse may substantially free up managers' time and efforts, thereby increasing their availability for other tasks. Its integrated, subject-oriented, time-variant, and non-volatile environment offers a stable and reliable information source for advanced organizational computing. Applications such as on-line analytical processing (OLAP) and data mining have been adopted to take advantage of the rich data source.

Despite the potential, success is not necessarily guaranteed. In fact, a survey revealed that more than 60 percent of companies owning a data warehouse rated the system as having only limited success in meeting user expectations [Stedman 1998]. The survey by Watson and Wixom [1997] showed that the ability of many existing data warehouses to provide users with easy and timely access to quality data is limited. Studies also indicated managerial or business customer resistance to data warehousing for several reasons including the fear of losing control, lack of technical skills, costs, and uncertainty about its utility [Ang and Teo 2000; Cooper et al. 2000]. Since the deployment of the data warehouse requires a multimillion-dollar investment and other resource mobilization, thorough evaluation of its values, risks, and return-on-investment should be conducted *a priori* [Moore and Wells 2000].

In spite of the increasing importance of the data warehouse, the lack of academic research limits our understanding of its individual and organizational implications [Watson and Wixom 1998; Wixom and Watson 2001]. Most of the existing literature consists of anecdotal reports about programs or quotations based on hearsay [Sakaguchi and Frolick 1996]. More systematic study is necessary to gain an in-depth understanding of the business implications and to justify future system investment. Given this background, an exploratory study was conducted to enhance general understanding of success factors in data warehousing. The focus of the study was on exploring the effect of variables pertaining to system quality, information quality, and service quality on user satisfaction for the data warehouse. Additional characterization was made of data warehouse users, their organizational tasks, and their usage of the data warehouse. The information systems success model [DeLone and McLean 1992, 2002] provided the overall conceptual framework for identifying the dimensions studied and corresponding variables.

Data were gathered from three different information sources. A survey was designed to collect relevant information from data warehouse users at a large enterprise. The survey data were analyzed through descriptive statistics and a multiple regression. Unstructured group interviews with end-users were performed to supplement the survey. Finally, frequent informal interviews during a two-month period with the information technology manager responsible for the data warehouse operation

provided additional background information including the architecture of the data warehouse, its usage, various benefits, and future expansion plans.

II. CHALLENGES OF DATA WAREHOUSING

Business organizations are facing stiff competition and increased uncertainties. To respond to such a turbulent business environment, an organization should be knowledge (or information)-based and an effective learning unit [Marquardt 1996]. Knowledge facilitates the efforts of an organization to reengineer itself through increased competitiveness, efficiency, and effectiveness. Any organization equipped with systematic procedures that utilize available knowledge is certainly in a position to outshine its competitors.

To be knowledge-based, a firm must be ready to detect, collect, process, and use available information effectively. Exponential growth of database volume offers organizations opportunities as well as challenges. Meanwhile, the rapid development of information technology has resulted in various tools that companies can adopt to increase their business intelligence and knowledge-processing capabilities. The data warehouse has surfaced as a key source of information for knowledge workers and managers. Its well-publicized value in offering high query-response performance and increased information accessibility, as well as being an integrated source of data, is creating an extremely popular environment for decision support at organizations [Watson and Wixom 1997].

The data warehouse is significnatly different from a conventional operational or transactional database in several aspects. First of all, a complex data structure must be maintained in order to offer flexible and dynamic retrieval of rich decision-support knowledge. For this, it maintains a data architecture that is more integrated, subject-oriented, time-variant, and non-volatile than transactional or operational databases [Inmon 1992]. Data structures should also be more cross-functional than "stovepipe" [Gardner 1998]. Realizing the performance and other functional needs requires system design and implementation approaches different from those of traditional databases. These include star-schema design, de-normalization, fact and dimensional tables, and different physical data structures (e.g., indexing). Data warehousing also has to deal with other issues of data management to facilitate effective decision support, which includes data quality insurance, the maintenance of historical data, and the production of derived data and attributes.

The data warehouse entails a high degree of complexity in the system's architecture because its data sources comprise transactional databases, operational data stores (ODS),[1] purchased data, and external data [Srivastava and Chen 1999; Wixom and Watson 2001]. The data warehouse architecture is a formal specification of data structure, communication, processing, and presentation of a data warehousing environment. Its architectural complexity is partially due to the various data sources and the sheer volume of data processed to adequately provide necessary information for knowledge workers, ad hoc power users, and standard report generators [Sen and Jacob 1998]. The system is expected to produce requested results in a reasonable amount of time despite possible processing delay caused by high computing demands. Improving the throughput and system response is one of the main challenges that system developers should address.

[1]The ODS uses data warehousing technology, but it is intended for assisting day-to-day operational activities rather than decision support-oriented analytical processing (Shin 2002).

Metadata development and management adds more complexity to the data warehouse. The semantic heterogeneity of data from multiple sources is one reason to maintain the metadata [Srivastava and Chen 1999]. This includes, but is not limited to, source system metadata (i.e., source system description schemas), data staging metadata (i.e., source target mapping information), DBMS metadata (i.e., view definition and indexes), and business logic and application supportive metadata (i.e., guidance for data elements, tables, and views). Metadata should be woven into data warehousing processes to drive business computing [Kimball et al. 1998]. The complexities and far-reaching business implications of the data warehouse justify a more focused look at the system issues separately from those of transactional or operational databases.

III. RESEARCH QUESTIONS AND VARIABLES STUDIED

CONCEPTUAL FOUNDATION

The theoretical notion of information systems success became the basis of this exploratory work. In fact, numerous studies have investigated the concept from different angles including its measurement and representation [Pitt et al. 1995; Raghunathan and Raghunathan 1994; Segars and Grover 1998]; and the technological, end-user, and organizational variables associated with it [Bruwer 1984; Li 1997; Yap et al. 1992; Yoon et al. 1995]. Grover et al [1996] characterized the research activities in terms of four research streams: *criteria demonstration, measurement, criteria relationship,* and *antecedents of IS effectiveness*. These studies concur that information systems success is a multidimensional concept that involves system factors (i.e., ease of use and throughput), data factors (i.e., data quality), organizational factors (i.e., policy, training, and support), and user factors (i.e., user satisfaction, individual impact, and utility) [Jain 1997; Li 1997; Williams and Ramaprasad 1996].

To explore potential success factors of the data warehouse, DeLone and McLean's [2002] IS success model was used; this is the revised version of their original model proposed in 1992. Despite some controversy on the integrity of the model [Seddon 1997], it remains the most influential model in conducting research on information systems success factors and therefore became a conceptual framework for identifying influencing factors of data warehouse success. The model suggests that there are six main dimensions of information systems success: *system quality, information quality* (i.e., usefulness of information)*, service quality, use, user satisfaction,* and *net benefits* (i.e., individual and organizational impacts). These represent technical (information production related), semantic (information itself), and effectiveness levels (information usage and its impact) of an information system [Delone and McLean 1992]. *System quality, service quality,* and *information quality* are expected to affect both *use* and *user satisfaction*. In the meantime, *use* and *user satisfaction* become the antecedents of *net benefits*. The dimensions, therefore, cover the entire spectrum of information flows from original production, through consumption, and all the way to their influence on individual and organizational performance. Wixom and Watson's [2001] empirical work showed the particular importance of system quality (i.e., system reliability and data quality) in securing benefits from the data warehouse.

THE VARIABLES

Among many potential variables of system quality, this survey included system throughput, ease of use, ability to locate data, access authorization, and data quality, which were regarded as crucial for the success of the data warehouse. System throughput of the data warehouse represents the response (or turnaround) time in retrieving the information sought by end-users and therefore closely associated to its processing speed [Sen and Jacob 1998]. Compared to those of transactional databases, queries submitted to the data warehouse, in general, require extensive computing because of their decision support nature. Too much processing delay, however, may force users to abandon the data warehouse because it may not be the only information source and its usage may not be as mandatory as that of operational systems serving daily transactions. From this perspective, the degree of system use may constitute a good indicator for data warehouse success [DeLone and McLean 2002]. Accordingly, system throughput is expected to be a key success factor that drives user-centric warehousing [Glassey 1998]. With its crucial nature, technical suggestions such as adaptive query caching [Saharia and Babad 2000] have been made to relieve possible performance bottlenecks and improve system throughput.

Perceived ease of use has also been frequently visited as an important indicator for information systems acceptance by end-users [Adams et al. 1992; Davis 1989]. Given the complexity of data processing for decision support, the perception of a system's ease of use may significantly affect the level of its adoption by prospective users. This study, therefore, extended existing work by investigating the implication of this variable in the data warehouse context.

As for ability to locate data, the massive data structure of the data warehouse frequently makes location of the right data difficult for end-users. Naturally, it has been frequently emphasized that the systematic management of metadata and its tight integration with computing processes become crucial for the successful adoption and use of the data warehouse by prospective end-users [Gardner 1998; Gray and Watson 1997; Kimball et al. 1998]. Also, the level of data detail (or granularity) may affect the location of the right data for users. Difficulty in the data search may not only deter data warehouse adoption by prospective users, but also negatively affect user satisfaction with the system and system-generated information. Measurement of this variable was, therefore, made from three different angles: general perception of abillity to locate data, ability to locate meta data (data field definition), and the level of details in data definition.

Access authorization evaluated perceived convenience of data access by data warehouse users. The data warehouse contains comprehensive, cross-departmental, cross-functional, and cross-hierarchical information that may be more sensitive than transactional data. Accordingly, data access authorization, data ownership, and data sensitivity become serious and difficult issues, especially when an organization has to deal with them at the global level [Sammon and Finnegan 2000]. Management of sensitive information in the enterprise-level system requires an adequate policy on access authorization. On the other hand, unnecessary or inadequate restriction of data access may hamper effective use of the database as a valuable knowledge source.

As another component of system quality, the critical importance of data quality in the data warehouse is well documented by existing work [Sakaguchi and Frolick 1996; Wixom and Watson 2001]. Four data quality variables (currency, accuracy, consistency, and level of detail) are included in the survey. Data currency (or recency), accuracy, and consistency are considered some of the most important attributes of data quality [Fox et al. 1994; Huh et al. 1990; Wang and Strong 1996]. Level of details (or granularity) reflects another quality feature of data representation [Fox et al. 1994].

Supporting adequate data granularity may be especially important in the data warehouse because of its role in facilitating rich knowledge extraction through flexible data search [Gray and Watson 1997; Kimball et al. 1998].

The *utility of information* acquired from the data warehouse represents a variable for the information quality dimension. Information utility of a system is its capability to support end-users to satisfy their information requirements [DeLone and McLean 1992]. This variable is, therefore, closely associated with the perceived usefulness of information obtained [Rivard and Huff 1985] and with the degree to which using a particular system could enhance the user's job performance [Davis 1989].

Many studies emphasized the importance of service quality for IS success [Barquin and Edelstein 1997; Bruwer 1984; Goodhue and Thompson 1995; Magal et al. 1988]. User training, as a representative *service quality* variable, has been repeatedly investigated as an effective way to attract potential users, enhance their understanding on the subject system, and increase user satisfaction [Chen et al. 2000; Rivard 1987]. Given the complexity of a data warehouse system and its data structure, end-user training could be especially crucial for its successful adoption and company-wide diffusion [Quaddus and Intrapairot 2001; Sakaguchi and Frolick 1996].

User satisfaction has been extensively adopted as a salient dependent variable to determine an information system's success [i.e., Delone and McLean 1992; McKeen *et. al*. 1994; Thong and Yap 1996]. In fact, user satisfaction is an important requirement in that it comprises one of the main considerations upon which managers can take corrective measures to increase system acceptance by end-users. Chen et al. [2000] suggested that the importance of user satisfaction also applies to the data warehouse. This variable was employed here to examine the relative impact of the selected independent variables on the data warehouse success. Table 1 summarizes the variables studied in each theoretical dimension.

Table 1. Variables Considered

IS Success Category	Variables Studied
System quality	• System throughput • Ease of use • Ability ot locate data • Access authorization • Data quality – Recency (currency) – Level of detail (abstractions) – Accuracy – Consistency
Information quality	• Information utility (usefulness)
Service quality	• User training
User satisfaction	• User satisfaction

IV. THE COMPANY AND DATA WAREHOUSING

The company studied is a major Fortune 500 enterprise with 65,000 employees and is one of North America's leading transportation, computer technology, and logistics companies. Its data warehousing effort was started in 1992 to increase operating efficiency, reduce expenses, and improve customer service. At the time, the company was keeping a tremendous amount of data on mainframe systems, all performing different functions and storing information in various formats. It was expected that by consolidating hundreds of departmental databases, much of the operational inefficiency resulting from redundant and often-conflicting information could be avoided. Also, the integrated system was expected to provide answers to critical business questions significantly faster and more reliably.

With full management support, an evaluation team of business professionals and research and development representatives from the company started a pilot data warehouse project. An accounts payable application was chosen for the pilot because it was small and fairly easy to manage, the data were clearly of value to multiple business areas, and the expected ROI (return on investment) was substantial and could easily be measured. The pilot system was estimated to save a $500,000 the first year, with a projected $2 million savings over four years. This ROI was a good enough reason to start a full-scale data warehouse project later on.

Teradata RDBMS on NCR 3600 was initially introduced for the platform, which was later upgraded to the NCR 5100 massively parallel system (MPP) that offers higher scalability and availability. Three 5100Ms were installed, one for testing, one for development, and one for production. At the time of the survey, the latter housed more than one terabyte of data in the subject areas of accounts payable, equipment maintenance, marketing, and movement information on transportation vehicles. Data stored in 1,500 relational tables were periodically updated from various on-line transactional and operational databases. It had been accessed by more than 2,400 users and usage was growing by about 100 people per month. At the time of the survey, most end-users were relying on Access DBMS from Microsoft as the front-end data access tool and a relatively small percentage of people were using Focus, a mainframe-based SQL interface.

V. RESEARCH METHODOLOGY

INTERVIEWS AND SURVEY DESIGN

Both interviews and a survey were utilized to gather relevant data. Two formal interviews, each lasting about two hours, were conducted as a form of group discussion. Managers and business customers from different functional departments provided input. In addition, an informal meeting was held on a weekly basis with a senior manager who was in charge of the data warehouse system. The informal discussions were highly informative and were held for about two months while the author was serving as a faculty intern at the company during a summer period. Through the interviews, information such as the data warehouse architecture, system specifications, financial effects, system usage, and future plans was obtained. Interviews also revealed detailed values of the data warehouse as well as roadblocks to its effective use.

The survey was designed to gather relevant information from existing data warehouse users (Appendix A). Seven-point Likert scales were used for many of the survey items. The first part of the survey obtained demographic information about data warehouse users including job title, position, and task characteristics. General information regarding current usage of the data warehouse was gathered in the next section, which contained items on access frequency, reasons for data warehouse access, and the types of analysis conducted. Questions on *system quality*, *information quality*, *service quality,* and *user satisfaction* with the system and with the information produced were included in the final section of the survey.

Most question items of the survey were developed to reflect the data warehouse environment. Four MIS professors assisted in the validation of question items in terms of their appropriateness and clarity. The database manager in charge of the data warehouse system also reviewed the survey questions from a practitioner's perspective to ensure they made sense to end users. The survey went through this iterative validation processe before it was distributed to the target users of the data warehouse.

SURVEY DATA COLLECTION

Data were collected from randomly selected data warehouse users from the company's 14 functional departments including finance, marketing, operations, and information technology (IT). There were five layers of employment status at the company: union workers, first-line (or lower) management, middle management, top management, and contracted consultants. It was confirmed by the IT manager that union workers and top management did not access the data warehouse. This was in line with the survey result from Watson and Wixom [1997], which showed that senior executive officers were not directly exploiting the data warehouse as end-users. Senior managers, although not system users themselves, were the main consumers of the reports produced from the data warehouse. Union workers were not responsible for generating decision support information and reports from the data warehouse and therefore were excluded from the sampling pool. Many respondents represented first-line and middle management. In addition, a significant number of contracted consultants were included in the selected sample. They were utilizing the data warehouse to undertake necessary services to the company. Heavy usage of the system by first-line and middle managers became the initial indication of the importance of the data warehouse in supporting managerial tasks.

The company's Lotus Notes network was used to distribute the electronic survey to 185 randomly chosen people and to gather their responses. A technical consultant from the company assisted the administration process. An additional solicitation was sent to non-respondents to increase the sample size and to avoid a potential non-response bias. The response rate was 35 percent with 65 respondents. The IT manager attributed the relatively low response rate to the on-going system integration project as a result of recent acquisition of a large transportation company. The newly acquired company depended heavily on outsourcing for the maintenance and upgrade of its IT/IS infrastructure. Naturally, the lack of internal know-how became a major problem in undertaking the large-scale system migration and integration. Table 2 summarizes the distribution of survey respondents in terms of their organizational status.

Table 2. Distribution of Respondents by Status

Organizational Status	Sample Size
First-line management	20
Middle management	27
Contracted consultants	17
Total	64

VI. SURVEY DATA ANALYSIS

Survey data were analyzed through two different statistical methods: descriptive statistics and regression analysis. The former was used for the general characterization of studied variables. The latter was applied to study the effect of system quality, information quality, and service quality variables on user satisfaction, exploring the proposed relationship of IS success factors [DeLone and McLean 2002] in the data warehouse context.

Analysis and interpretation of data that use the seven-point Likert scale can be highly subjective in nature and the potential bias may further increase when the scale is numerically coded (i.e., from 1 to 7) as in this case. To maintain consistency in the interpretation and discussion of statistical results, therefore, the following heuristics are used: very low to low (1-3), mid to mid-high (3.1-5), and high to very high (5.1-7).

DESCRIPTIVE STATISTICS

Task Characterization

Implications of task characteristics on IS usage have been frequently investigated. For instance, empirical studies demonstrated that task complexity is a contingency factor that could significantly affect IS usage by prospective users [Shin et al. 1999; Zigurs and Buckland 1998]. Users of the data warehouse were characterized by the nature of their organizational tasks. Average values indicated that data warehouse users, in general, dealt with complex tasks (5.57/7) that required frequent decision making (5.51/7). Many respondents agreed that their tasks were cross-departmental in nature (5.54/7) and that they needed extensive access to company information (6.09/7) to undertake assigned tasks. Overall, respondents were neutral about the problem of information overload (3.51/7). However, certain users appeared to suffer from lack of information (4.68/7) and low reliability of available information (4.68/7) from the data warehouse.

Next, data analysis was limited to a comparison of the responses between first-line management and middle management (Table 3). Two sample t-tests indicate that middle management dealt with more complex (Q2) and cross-functional (Q4) tasks than first-line management. Overall, the chance that managers' tasks were being hampered by information overload appeared to be low, especially for first-line management. The average response, on the other hand, indicated that too much information could pose a problem for effective decision making by middle managers (Q7).

Table 3. Comparison between First-Line Management and Middle Management

Question Items	First-Line Management		Middle Management		T-Ratio	P-Value
	Mean	SD	Mean	SD		
Tasks nature						
Q1. Task repetitveness	3.50	1.70	4.18	1.52	-1.45	.153
Q2. Tasks complexity	5.20	.95	5.85	.81	-2.52	.015*
Q3. Frequent decision making	5.45	1.09	5.85	.94	-1.34	.186
Q4. Cross-functional tasks	5.10	1.29	6.03	1.01	-2.77	.008*
Information and decision-making						
Q5. Extensive information use	5.95	.88	6.29	.86	-1.33	.187
Q6. Lack of information	4.55	1.50	4.74	1.76	-.38	.699
Q7. Too much information	2.95	1.63	4.22	1.71	-2.56	.014*
Q8. Lack of information reliability	4.25	1.97	4.92	1.79	-1.22	.227

*Equal variances are assumed.

Usage of Data Warehouse

Average frequency of data warehouse access was about 15 times per day (see Table 4 for the frequency distribution). Most users (69 percent) accessed the system less than 10 times per day. Ten accesses per day was the mode (12 counts) of the frequency distribution. Two users used the data warehouse constantly during the day, as indicated by 100 accesses per day. Average frequencies were 10.3 and 15.1 for first-line management and middle management, respectively. Middle managers appeared to use the data warehouse more frequently, but two-sample t-tests did not support statistical significance.

Table 4. Frequency Distribution of Data Warehouse Access

Access per Day	Frequency
≤ 10	43
11–20	9
21–30	3
31–40	3
41–100	4

Types of tasks for which the data warehouse was used were investigated. For this, eight main organizational tasks were included in the survey (Table 5). The first four (decision-making support, status monitoring, planning, and forecasting) were considered more unstructured than the others.

The data gathered were then summarized in terms of user access ratio for each task and daily access frequency.

In general, more users were using the data warehouse for unstructured duties rather than for routine or administrative responsibilities. For instance, more than 70 percent of respondents were taking advantage of the data warehouse for decision support. The average frequency of daily use was highest for decision making (3.45) and status monitoring (2.82) followed by administration (2.18) and accounting (1.47). Usage for human resource and labor relationship management was reported as minimal.

Table 5. Usage of the Data Warehouse

Types of Work	Access		Access per Day	Std.
	Yes	No		
Decision-making support	45 (70.3%)	19	3.45	7.29
Status monitoring	38 (59.4%)	26	2.82	9.02
Planning	28 (43.7%)	36	0.73	1.98
Forecasting	20 (31.2%)	44	0.63	1.99
Administration	19 (29.7%)	45	2.18	12.6
Accounting	19 (29.7%)	45	1.47	4.59
Resource allocation/budgeting	13 (20.3$)	51	0.33	1.27
Human resource/labor relations	3 (7.8%)	59	0.02	0.13
Other	12 (18.7%)	52	0.98	3.29

End-User Perceptions

While respondents were on the whole positive about data quality, perceived differences appeared to exist among the quality dimensions (Table 6). Data currency and the level of details received relatively high acceptance from users. Data consistency received the lowest acceptance. This indicates that, in this particular case, the most challenging aspect of data warehousing was in securing an adequate level of data consistency in the system.

When first-line and middle management responses were compared, only the level of details showed a significant difference. Middle management was more positive than first-line management (4.5/7) that the data in the data warehouse maintained the level of details needed for their tasks (5.66/7) ($p = 0.004$). This result appears to confirm that first-line managers needed more detailed information than did middle managers to conduct their duties. It became an indication that the information needs of system users differed in the level of abstraction and structure, and that the success of the data warehouse partially depended on its flexibility to satisfy heterogeneous demands from information seekers at different levels.

Data summarization revealed that an ability to locate data problem could be a major roadblock to the effective use of the data warehouse. Users responded that data location was not very easy (3.7/7) and the definition of data attributes available was not detailed enough to support their tasks (3.4/7). The data warehouse was, in general, recognized as being critical to improving the productivity

of knowledge workers (5.9/7). During the interviews it was learned that, despite the system's importance, reduced throughput and slow system response could significantly hamper user productivity. It appeared that the slow response was caused partially by heavy hits to the data warehouse (at least 10,000 queries per day), by the data structure not reflecting user views, and by the misusage (e.g., ineffective queries) of the system from certain users. Business users suggested that the data structure be task-oriented and structured to minimize the number of queries necessary to recover a piece of information.

Table 6. Perceptions on Data Quality

Data Quality	Average
Data currency	4.9/7
Level of details	4.9/7
Data consistency	4.1/7
Accuracy, reliability	4.6/7

REGRESSION ANALYSIS

Independent Variables

A regression analysis was conducted to evaluate the strength of association between variables of system quality, information quality, and service quality, and user satisfaction for the data warehouse. For this, the survey gathered information on user satisfaction in terms of the overall quality of information generated (Q22) and the overall performance of the data warehouse system (Q23). Two items showed high factor loading and inter-item reliability (Cronbach alpha = .751) and the overall satisfaction level was computed by averaging their standardized scores (see Appendix B). Independent variables were composed of data quality (Q14-Q17), ability to locate data (Q18-Q20), access authorization (Q21), ease of use (Q24), user training (Q25), system performance (Q26-Q27), and information utility (Q28-Q29). It was expected that high quality of data, ease of locating relevant data, ease of use for the system, increased system throughput, less-restrictive data accessibility, adequate user training, and increased utility of recovered information would be positively associated with user satisfaction with the data warehouse [Delone and McLean 1992; Goodhue and Thompson 1995].

Preprocessing of Data

Preprocessing of survey data was conducted prior to the analysis. First, convergence tests for variables with multiple indicator items were conducted through confirmatory factor analysis (see Appendix B). Indicators belonging to a construct, in general, showed a statistical significance in their convergent validity. An indicator (Q17) from the data quality construct had a low factor loading and was dropped from further consideration. Second, aggregations for the multi-indicator items were

conducted. To ensure the equal contribution of weights from each indicator, the raw data were first standardized with Z-scores. Then the standardized values were averaged to get the unit-weight scores for a multi-indicator construct. When there was only one indicator, the raw data were standardized with Z-scores. A multiple regression test was then conducted using the standardized scores.

Analysis Results

Data quality, ability to locate data, and system throughput explained significant variations in user satisfaction (Table 7). In fact, these variables were frequently blamed by interviewees as major constraints of effective data warehouse usage. The coefficients confirmed that higher data quality, easy data location, and improved system throughput were significantly associated with increased user satisfaction for the data warehouse. Care should be given in interpreting the result, though. It does not mean that the other variables (accessibility, ease of use, user training, and information utility) are not important for the data warehouse system. It merely shows which variables had a stronger association with the satisfaction of end-users.

Table 7. Initial Estimation of the Regression Model

	Beta	S.E.	t	p-level
Intercept			.000	1.000
Data quality	.339	.099	3.560	.001*
Ability to locate data	.352	.095	3.643	.001*
Access authorization	-.058	.078	-.655	.515
Ease of use	.099	.089	1.002	.321
User training	-.013	.071	-.166	.868
System throughput	.287	.090	3.293	.002*
Information utility	.022	.081	.269	.789

Next, parameter estimation based on a more parsimonious model was conducted to compare the explanatory power of the significant variables (Table 8). The ANOVA table confirmed statistical significance of the parsimonious model ($p = 0.000$). R^2 (0.700) indicated that 70 percent of variations in user satisfaction was explained by the three variables. As observations were standardized, comparison of the Beta coefficients indicated relative influences of the variables on user satisfaction. Ability to lcoate data showed the highest association with the dependent variable followed by data quality and system throughput, although the difference was small.

Independent variables were checked for multicollinearity, which can bias parameter estimation. According to the rule of thumb, if a correlation coefficient $|r_{xy}|$ is greater than 0.7, multicollinearity becomes a substantial threat [Anderson et al. 1996]. The Spearman correlation matrix indicated a certain degree of correlation among independent variables (see Appendix C), but it wasn't high enough to cause multicollinearity and bias the test results.

Table 8. Regression Equation and Anova Tables

		Beta	S.E.	t	p-level
Data quality		.366	.087	4.19	.000
Ability to locate		.377	.086	4.36	.000
Throughput		.286	.083	3.43	.001
	Sums of Squares	**df**	**Mean Squares**	**F**	**p-level**
Regression	35.386	3	11.79	46.64	.000
Residual	15.171	60	.252		
Total	50.557	63			

VII. DISCUSSION

SYNTHESIS

The analysis based on the survey data revealed factors that led the data warehouse toward becoming a more productive environment. First-line and middle managers and other contracted knowledge workers took advantage of the system to support unstructured as well as routine decision making and to seek information necessary for their work. Average daily access frequency (15 times) rendered a strong indication that information recovered from the data warehouse was vital for the increased work productivity of many knowledge workers. Although it was adopted for a variety of tasks from routine administration to external customer services, more users took advantage of it for advanced data analysis and decision support. This confirms Cooper et al.'s [2000] study, where the utility of the data warehouse was better realized in solving unstructured problems such as customer strategy development than in using it for routine duties. The analysis revealed that information demand from system users varied in the level of abstractions and information structures, and that the success of the data warehouse could partially depend on its flexibility in satisfying heterogeneous requirements from information seekers.

The study of user perceptions of the data warehouse uncovered areas that warrant more attention if it is to be effectively adopted by an organization. In spite of the fact that the system has become a classic data warehousing project success story, the study indicated that low data quality could still be problematic to end-users as pointed out by several studies [Ballou and Tayi 1999; Watson and Wixom 1997]. The regression analysis indicates that data quality can affect system users' satisfaction significantly. Among the data quality dimensions considered, the lack of data consistency (i.e., field naming, data duplication, and data fragmentation) appeared to be especially troublesome to system users. Data inconsistency is, in general, a difficult problem to address because the data warehouse imports data from various sources that have evolved independently without much regard to data quality [Sammon and Finnegan 2000]. Accordingly, much attention should be placed on correcting data inconsistency before and during the data staging process [Sammon and Finnegan 2000]. If the data in the data warehouse are imported from transactional systems as a form of data dump without much transformation and quality validation, most of the

quality problems in the transactional database will simply be inherited. The data quality issue is important because the data become non-volatile once they are accommodated in the data warehouse [Inmon 1992].

Interviews also revealed that there could be differences among data warehouse stakeholders regarding the emphasis of data quality dimensions. For instance, those responsible for the operation and management of the data warehouse gave data the highest priority. Meanwhile, getting timely information into the hands of decision makers was more crucial for end-users. Although it is not surprising that stakeholders had varied views on the importance of data quality dimensions, this may become an indication that such differences in the priority of data characteristics have to be effectively incorporated and balanced in the data warehouse design stage.

The regression analysis pointed out that slow system response substantially affected knowledge workers' satisfaction with the data warehouse. This was the case even though the data warehouse was housed on the scalable parallel machine designed to handle heavy computing demand. In the meantime, interviews revealed that management issues such as the misuse of the data warehouse, the lack of user training and subsequent system abuse, and inefficient data modeling that didn't reflect end-user perspectives significantly contributed to the degradation of system performance. The people interviewed emphasized the special importance of end-user training not only to speed up the system adoption [Quaddus and Intrapairot 2001], but also to reduce computing errors or inefficiency that could trigger a significant penalty in the system response. The paucity of *deliberate intelligence* in database queries submitted by end users sometimes wasted costly hours of computer processing time, reaffirming the critical importance of user competence in data warehouse computing. In addition, the examination and design of data structures from the perspective of current usage and user needs was said to be important to improve system throughput.

The regression analysis also showed significant impact of ability to locate data on user satisfaction for the data warehouse. Survey analysis and interviews consistently exposed that the difficulty of data location was a major hurdle for the effective use of the data warehouse. The group interviews revealed contributing factors of the difficulty in data location. First, the lack of user knowledge regarding the data structure of the data warehouse contributed to the difficulty of locating the right information by end users. Offering more education on the information structure of the data warehouse was thought to relieve the problem. Second, there was a lack of comprehensive multi-level meta-data readily available for end users. For instance, users pointed out that, although the system allowed recording of a short descriptive comment for each table, view, and column, many comments did not provide enough contextual information for them to be able to understand the data semantics. Accordingly, efforts to improve end-user documentation to provide rich meta-data and to adopt an advanced methodology such as active meta-data [Kimball et al. 1998] to tightly integrate them with database processing were thought crucial to addressing this problem. Third, it was learned that, depending on the type of difficulty encountered during system use, end users should be given a specific contact point to seek assistance. The absence of designated staff to provide relevant advice appears to have aggravated the ability to locate data problem.

Finally, it should be stressed that the availability of data access tools could affect user perceptions in various manners. A managed query environment could make it easier than writing SQL queries to locate data, making the underlying data model largely transparent to users. Availability of easy-to-use data access tools may, therefore, significantly affect users' satisfaction with the data warehouse. For instance, in the case studied, people were relying mostly on Access and sometimes mainframe-based Focus as the query tool. They appeared to be more satisfied with user-friendly Access than

Focus. For instance, one user said "Excellent [indicating Access], far better than Focus." Meanwhile, another user pointed out a potential weakness of Access as follows: "It is great for building small applications, but when testing these applications before they are up and running in Production, it is not exactly the best query testing tool. It locks up the entire workstation with each query even while running."

The problems brought up during the interviews largely coincided with the results of the survey analysis. They were primarily related to system quality and service quality issues, and major concerns of system quality were associated with system design, accessibility, and the quality of data imported from source systems. These are summarized in Table 9.

Table 9. Problems Encountered during the Use of the Data Warehouse (Summary of Interviews)

Categories
System Quality • System design issues – Slow response – Too many steps to get information – Poor data modeling – Lack of audit trails such as the time of updat – Lack of meta-data • Accessibility – Limited acccess to the data warehouse • Data quality issues – Lack of recency – Low data accuracy and missing data – Data format anomaly – Duplicate records – Inconsistent field names – Low data reliability and consistency
Service Quality – Lack of appropriate user training – Lack of contact information

ADDITIONAL PERSPECTIVES

User interviews offered additional perspectives on the benefits of the data warehouse at the company, although they are closely related to the unique position of the company. Participants agreed that the return on investment for the data warehouse was well justified through considerable gains in productivity and enhanced quality of customer service. They indicated that the system helped the company in a variety of ways. Decision making and other organizational operations became more fact based, thereby improving the quality and effectiveness of decision-making processes and outcomes. Fact-based operation facilitated distributed decision making by more people. The

distribution of decision making, in turn, improved organizational efficiency by reducing procedural overhead and quickly accommodating fluctuations in the marketplace. Information from the data warehouse created proactive rather than reactive and dynamic rather than static processes, and allowed flexible responses to cross-functional operations that were prone to change. Knowledge from the data warehouse also enabled market-oriented integration of activities in such areas as forecasting, operation planning, customer analysis, and asset utilization.

It was learned that the management of the data warehouse took more than 20 full-time-equivalent staff. Nonetheless, there was no doubt that the return far exceeded the cost. Some of the benefits could be quantified. It was estimated that an annual savings of $3.5 million was achieved by better vehicle management and by cutting the accounts receivable cycle. In fact, since 1992 the company has spent an estimated $20 million for data warehouse projects, which included hardware, systems integration work, consultants' fees, and staff and end-user training. In spite of the technical and financial challenges, there had been enough payoffs. Buoyed by the data warehouse success, the company was looking into a more advanced form of computing, especially data mining, which attempts to recover hidden but valuable knowledge from the database source. The company learned that being equipped with systematic measures for recovering and utilizing information from the data warehouse could place it in a better competitive position.

RESEARCH LIMITATIONS

This research has limitations in the research methodology used. First, this study has strong attributes of a case study because the data were gathered from employees of a single company. Accordingly, certain system or non-system issues discussed here might be phenomena local to the studied organization. In that sense, this is an intensive rather than an extensive study in which the generalizability of analysis results may be lacking. Meanwhile, the particular data warehouse system was recognized among industry watchers as a showcase for data warehousing success. Its system design and implementation and its successful adoption by prospective users were, therefore, expected to offer a reliable setting from which research questions could be rigorously pursued. Also, that the survey data represented data warehouse users from 14 different functional departments might have a positive effect on the generalizability of the study.

A reduced number of indicator items in certain constructs might have affected their measurement reliability and accordingly the power of the regression analysis. For instance, two indicators were used for user satisfaction, system throughput and information utility, although the Cronbach alphas for all constructs (ranged from .639 to .897) indicated their reasonable stability. Three constructs (accessibility, ease of use, and user training) were also represented by a single item.

Limited sample size (64) might have affected the integrity of the statistical inference. For instance, a sample size of 64, does not meet the rule-of-thumb ratio of 1 item to 10 data points for stable factor loadings [Nunnally 1978]. The t-tests in Table 3 were also based on relatively small record sets of 20 for first-line management and 27 for middle management. Given that the survey was distributed within an organization, a 35 percent response rate could be regarded as rather low. Besides, although the survey instrument was reviewed and validated by five people from both academia and industry, more content validation could have been done through pilot testing the questionnaire items.

Finally, group interviews, as an additional information source, may pose a reliability problem as interviewees might not have been forthcoming with their opinions on sensitive issues. This might limit the usefulness of the case study in providing rich information to readers. Such risk of participants holding back opinions was expected to be relatively low for two reasons. First, participants were from different functional departments and little conflict of interest was expected in discussing problems and benefits associated with the data warehouse system. Second, most participants, as system users, were of similar ranks as shown in the survey, which could have resulted in more open discussions. In the meantime, I would like to emphasize the merit of adopting multiple methods of data collection when the target system is as complicated as this case because additional data sources such as repetitive interviews render valuable perspectives that are not conceived at the stage of survey design.

VIII. CONCLUDING REMARKS

The data warehouse is taking the corporate sector by storm [Sen and Jacob 1998]. It deals with a database that is orders of magnitude larger than a conventional database. It is subject to complex ad hoc as well as regular queries to mostly assist decision-support processes. Data warehousing poses various challenges such as migration of data from legacy systems, maintenance of data quality, system management, and user education and training. The main focus of this study was to empirically investigate the technological and non-technological factors that affected the satisfaction of data warehouse users through a survey and interviews. Additional investigation was conduced on data warehouse end users and their data warehouse usage, the value of the data warehouse in supporting managerial and knowledge workers' tasks. This study confirmed that the system was critical in serving a variety of tasks from routine work to complex planning and decision making. Users had a high regard for the role of the data warehouse in enhancing their work productivity. On the other hand, concerns that could adversely affect user satisfaction consistently surfaced from the survey analysis and during interviews. Many of the concerns were related to system and service quality, some of which could be addressed relatively easily via correctional efforts. In general, the study indicates that the IS success model introduced by DeLone and McLean [2002] can become a good framework in understanding the success of data warehousing. In the meantime, with its decision support focus and unique system characteristics (i.e., data structure and volume), critical success factors in the categories of information, system, and service quality could be different from those of more traditional information systems. For instance, ability to locate data and the lack of meta-data may not pose such a significant roadblock to end users in transactional information systems. Further investigation is necessary for improved understanding on this issue. Overall, the survey and interviews showed that the data warehouse as a decision support backbone could become a strategic tool to augment organizational competency.

Editor's Note: This paper was first received on November 11, 2002. The article was with the author two months for one revision. Phillip Ein-Dor was the editor.

IX. REFERENCES[2]

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

Anderson, D. R., Sweeney, D. J., and Williams, T. A. *Statistics for Business and Economics*, St. Paul, MN: West Publishing Company, 1996.

Ang, J., and Teo, T. S. H. "Management Issues in Data Warehousing: Insights from the Housing and Development Board," *Decision Support Systems* (29:1), 2000, pp. 11-20.

Ballou, D. P., and Tayi, G. K. "Enhancing Data Quality in Data Warehouse Environments," *Communications of the ACM* (42:1), 1999, pp. 73-78.

Barquin, R., and Edelstein, H. *Building, Using, and Managing the Data warehouse*, Upper Saddle River, NJ: Prentice Hall, 1997

Bruwer, P. J. S. "A Descriptive Model of Success for Computer-Based Information Systems," *Information & Management* (7:2), 1984, pp. 63-67.

Chen, L., Soliman, K. S., Mao, E., and Frolick, M. N. "Measuring User Satisfaction with Data Warehouses: An Exploratory Study, *Information & Management* (37:3), 2000, pp. 103-110.

Cooper, B. L., Watson, H. J., Wixom, B. H., and Goodhue, D. L. "Data Warehousing Supports Corporate Strategy at First American Corporation," *MIS Quarterly* (24:4), 2000, pp. 547-567.

Davis, F. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly* (13:4), 1989, pp. 319-340.

Delone, W. J., and McLean, E. R. "Information Systems Success: The Quest for the Dependent Variable," *Information Systems Research* (3:1), 1992, pp. 60-95.

Delone, W. J., and McLean, E. R. "Information Systems Success Revisited," in *Proceedings of the 35th Hawaii International Conference on System Sciences*, Los Alamitos, CA: IEEE Computer Society Press, 2002.

Fox, C., Levitin, A., and Redman, T. "The Notion of Data and Its Quality Dimensions," *Information Processing & Management* (30:1), 1994, pp. 9-19.

Gardner, S. R. "Building the Data Warehouse," *Communications of the ACM* (41:9), 1998, pp. 52-60.

Glassey, K. "Seducing the End User," *Communications of the ACM* (41:9), 1998, pp. 62-69.

Goodhue, D. L., and Thompson, R. L. "Task-Technology Fit and Individual Performance," *MIS Quarterly* (19:2), 1995, pp. 213-236.

Gray, P., and Watson, H. *Decision Support in the Data Warehouse,* Upper Saddle River, NJ: Prentice Hall, 1997.

Grover, V., Jeong, S. R., and Segars, A. H. "Information Systems Effectiveness: The Construct Space and Patterns of Application," *Information & Management* (31:4), 1996, pp. 177-191.

---

[2]Editor's Note: The following reference list contains the address of World Wide Web pages. Readers who have the ability to access the Web directly from their computer or are reading the paper on the Web can gain direct access to these references. Readers are warned, however, that

1. These links existed as of the date of publication but are not guaranteed to be working thereafter.
2. The contents of Web pages may change over time. Where version information is provided in the References, different versions may not contain the information or the conclusions references.
3. The authors of the Web pages, not JAIS, are responsible for the accuracy of their content.
4. The author of this article, not JAIS, is responsible for the accuracy of the URL and version information.

Huh, Y. U., Keller, F. R., Redman, T. C., and Watkins, A. R. "Data Quality," *Information and Software Technology* (32:8), 1990, pp. 559-565.

Inmon, W. H. *Building the Data warehouse*, Wellesley, MA: QED Technical Publishing Group, 1992.

Jain, R. "Key Constructs in Successful IS Implementation: South-East Asian Experience," *Omega* (25:3), 1997, pp. 267-284.

Kimball, R., Reeves, L., Ross, M., and Thornthwaite, W. *The Data Warehouse-Lifecycle Toolkit*, New York: Wiley, 1998.

Li, E. Y. "Perceived Importance of Information System Success Factors: A Meta Analysis of Group Differences," *Information & Management* (32:1), 1997, pp. 15-28.

Magal, S. R., Carr, H. H., and Watson, H. J. "Critical Success Factors for Information Center Managers," *MIS Quarterly* (12:3), 1988, pp. 413-425.

Marquardt, M. J. *Building the Learning Organization*, New York: McGraw-Hill, 1996.

McKeen, J. D., Guimaraes, T., and Wetherbe, J. C. "The Relationship between User Participation and User Satisfaction: An Investigation of Four Contingency Factors," *MIS Quarterly* (18:4), 1994, pp. 427-451.

Moore, A., and Wells, D. "How to Do a Data Warehouse Assessment (And Why)," Knowledge Partners, Inc., available online at **http://www.kpiusa.com/ReadingRoom/DWH%20Assessment.htm**; retrieved on November 13, 2000 (originally published in *Journal of Data Warehousing*).

Nunnally, J. C. *Psychometric Theory*, New York: McGraw-Hill, 1978.

Pitt, L. F., Watson, R. T., and Kavan, C. B. "Service Quality: A Measure of Information Systems Effectiveness," *MIS Quarterly* (19:2), 1995, p. 173-187.

Quaddus, M., and Intrapairot, A. "Management Policies and the Diffusion of Data Warehouse: A Case Study Using System Dynamics-Based Decision Support System," *Decision Support Systems* (31:2), 2001, pp. 223-240.

Raghunathan, B., and Raghunathan, T. S. "Adaptation of a Planning System Success Model to Information Systems Planning," *Information Systems Research* (5:3), 1994, pp. 326-340.

Rivard, S. "Successful Implementation of End-User Computing," *Interfaces* (17:3), 1987, pp. 25-33.

Rivard, S., and Huff, S. L. "An Empirical Study of Users as Application Developers," *Information & Management* (8:2), 1985, pp. 89-102.

Saharia, A., and Babad, Y. M. "Enhancing Data Warehouse Performance through Query Caching," *Database for Advances in Information Systems* (31:3), 2000, pp. 43-63.

Sakaguchi, T. and Frolick, M. N. "A Review of the Data Warehousing Literature," Unpublished Manuscript, 1996, available online at **http://www.nku.edu/~sakaguch/dw-web.htm** (retrieved on November 13, 2000).

Sammon, D., and Finnegan, P. "The Ten Commandments of Data Warehousing," *Database for Advances in Information Systems* (31:4), 2000, pp. 82-91.

Seddon, P. B. "A Respecification and Extension of the DeLone and McLean Model of IS Success," *Information Systems Research* (8:3), 1997, pp. 240-253.

Segars, A. H., and Grover, V. "Strategic Information Systems Planning Success: An Investigation of the Construct and Its Measurement," *MIS Quarterly* (22:2), 1998, pp. 139-163.

Sen, A., and Jacob, V. S. "Industrial-Strength Data Warehousing," *Communications of the ACM* (41:9), 1998, pp. 28-31.

Shin, B. S. "A Case of Data Warehousing Project Management," *Information & Management* (39:7), 2002, pp. 581-592.

Shin, B. S., Higa, K., Sheng, O., and Ide, T. "Analyzing the Media Usage Behavior of Telework Groups: A Contingency Approach," *IEEE Transactions on Systems, Man, and Cybernetics* (29:1), 1999, pp. 127-139.

Srivastava, J., and Chen, P. Y. "Warehouse Creation—A Potential Roadblock to Data Warehousing," *IEEE Transactions on Knowledge and Data Engineering* (11:1), 1999, pp. 118-126.

Stedman, C. "Warehousing Projects Hard to Finish," *Computerworld* (32:12), 1998, p. 29.

Thong, J., and Yap, C. "Information Systems Effectiveness: A User Satisfaction Approach," *Information Processing and Management* (32:5), (5), 1996, pp. 601-610.

Wang, R. Y., and Strong, D. M. "Beyond Accuracy: What Data Qquality Means to Data Customers," *Journal of Management Information Systems* (12:4), 1996, pp. 5-34.

Watson, H. J., and Wixom, B. H. "Data Warehousing: A Framework and Survey of Current Practices," *Journal of Data Warehousing* (2:1), 1997, pp. 10-17.

Watson, H. J., and Wixom, B. H. "Managerial Considerations," *Communications of the ACM* (41:9), 1998, pp. 32-37.

Williams, J. J., and Ramaprasad, A. "A Taxonomy of Critical Success Factors," *European Journal of Information Systems* (5:4), 1996, pp. 250-260.

Wixom, B. H., and Watson, H. J. "An Empirical Investigation of the Factors Affecting Data Warehousing Success," *MIS Quarterly* (25:1), 2001, pp. 17-41.

Yap, C. S., Soh, C. P. P, and Raman, K. S. "Information System Success Factors in Small Business," *Omega* (20:5/6), 1992, pp. 597-609.

Yoon, Y., Guimares, T., and ONeal, Q. "Exploring the Factors Associated with Expert Systems Success," *MIS Quarterly* (19:1), 1995, pp. 83-106.

Zigurs, I., and Buckland, B. K. "A Theory of Task/Technology Fit and Group Support Systems Effectiveness," *MIS Quarterly* (22:3), 1998, pp. 313-334.

X. ABOUT THE AUTHOR

Bongsik Shin is an associate professor at the Department of Information and Decision Systems at San Diego State University. He earned a Ph.D. from the University of Arizona. Bongsik was on the faculty of Management Information Systems at the University of Nebraska, Omaha, from 1997 through 1999. His work has been published or accepted for publication in *Communications of the ACM, IEEE Transactions on Systems, Man, and Cybernetics, IEEE Transactions on Engineering Management, Journal of Data Warehousing, Journal of Organizational Computing and Electronic Commerce, Journal of Systems Integration, Journal of Management Information Systems, Information & Management,* and *Journal of Database Management.* He has been teaching telecommunications and networking, electronic commerce, and data warehousing.

Appendix A
Survey on the Use of Data Warehouse

Preview question: What is a single most challenging task in your job (Please be as specific as you can)?

Answer:

I. Personal Information

Department:
Job Title:
My position at X-company fall under (Check one)

a. Management - Grade 12-17
b. Management - Grade 18-23
c. Management - Grade 24>
d. Others

If I assess my work:

1. The tasks I deal with are repetitive in nature.
   Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

2. The tasks I deal with are complex.
   Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

3. The tasks I deal with involve frequent decision-making.
   Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

4. The tasks I deal with frequently involve more than one business function.
   Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

5. The tasks I deal with require extensive access to company information
   Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

6. The quality of my decision-making suffers due to the lack of information I need.
   Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

7. Frequently, too much information distracts me in my decision-making.
   Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

8. The quality of my decision-making suffers because available information lacks reliability.
   Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

II. Use of Existing Data Warehouse

9. How frequently do you access the data warehouse?
   [ ] times a day, or
   [ ] times a week, or
   [ ] times a month, or
   [ ] do not access (Please go to question items on data-mining in section III)

10. How much time, on average, do you spend on each access to the data warehouse?
    [ ] minutes

11. For what purposes, do you access the data warehouse?

    11a. To create reports for supervisors or managers: Yes[ ], No[ ]
    If yes, how frequent?
    [ ] times a day/ [ ] times a week/ or [ ] times a month

    11b To acquire information for my own work: Yes[ ], No[ ]
    If yes, how frequent?
    [ ] times a day/ [ ] times a week/ or [ ] times a month

    11c. For external customer services: Yes[ ], No[ ]
    If yes, how frequent?
    [ ] times a day/ [ ] times a week/ or [ ] times a month

    Please describe the customer services:

    11d. Other tasks Yes[ ], No[ ]
    If yes, how frequent?
    [ ] times a day/ [ ] times a week/ or [ ] times a month

    Please describe the tasks:

12. If you create reports for supervisors or managers, what are the types and the frequency of the report (Please skip these questions if you do not generate reports)?

    Types of report generated

    12a Change analysis: Yes[ ], No[ ]
    If yes, how frequent?
    [ ] times a day/ [ ] times a week/ or [ ] times a month

    Please describe types of analysis:

12b Summarization: Yes[ ], No[ ]
If yes, how frequent?
[ ] times a day/ [ ] times a week/ or [ ] times a month

Please describe types of analysis:

12c Trend analysis: Yes[ ], No[ ]
If yes, how frequent?
[ ] times a day/ [ ] times a week/ or [ ] times a month

Please describe types of analysis:

12d Measure comparison: Yes[ ], No[ ]
If yes, how frequent?
[ ] times a day/ [ ] times a week/ or [ ] times a month

Please describe types of analysis:

12e Others: Yes[ ], No[ ]
If yes, how frequent?
[ ] times a day/ [ ] times a week/ or [ ] times a month

Please describe types of analysis:

13. If you access the data warehouse *for your own work*, what are the tasks and the frequency of access?

13a Status monitoring Yes[ ], No[ ]
If yes, how frequent?
[ ] times a day/ [ ] times a week/ or [ ] times a month

Please describe types of analysis:

13b Decision-making support Yes[ ], No[ ]
If yes, how frequent?
[ ] times a day/ [ ] times a week/ or [ ] times a month

Please describe types of analysis:

13c Forecasting Yes[ ], No[ ]
If yes, how frequent?
[ ] times a day/ [ ] times a week/ or [ ] times a month

Please describe types of analysis:

13d Administration Yes[ ], No[ ]
If yes, how frequent?
[ ] times a day/ [ ] times a week/ or [ ] times a month

Please describe types of analysis:

13e Accounting Yes[ ], No[ ]
If yes, how frequent?
[ ] times a day/ [ ] times a week/ or [ ] times a month

Please describe types of analysis:

13f Planning Yes[ ], No[ ]
If yes, how frequent?
[ ] times a day/ [ ] times a week/ or [ ] times a month

Please describe types of analysis:

13g Resource allocation, budgeting Yes[ ], No[ ]
If yes, how frequent?
[ ] times a day/ [ ] times a week/ or [ ] times a month

Please describe types of analysis:

13h Personnel management, labor relations Yes[ ], No[ ]
If yes, how frequent?
[ ] times a day/ [ ] times a week/ or [ ] times a month

Please describe types of analysis:

13i Others Yes[ ], No[ ]
If yes, how frequent?
[ ] times a day/ [ ] times a week/ or [ ] times a month

Please describe types of analysis:

III. Perception of Information from the Data Warehouse

Please provide your perceptions of the existing data warehouse system.

Data Quality

14. I can get data that is current enough to meet my work needs.
Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

15. The data warehouse maintains data at an appropriate level of detail for my tasks.
    Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

16. The data in the data warehouse is accurate and reliable.
    Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

17. There are times when I find that supposedly equivalent data is inconsistent in the data warehouse.
    Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

Ability to Locate Data

18. It is easy to locate data on a particular issue in the data warehouse, even if I haven't used that data before.
    Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

19. The exact definition of data fields relating to my tasks is easy to obtain.
    Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

20. Definition of data in the data warehouse provides the level of detail that satisfies my requirements.
    Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

Access Authorization

21. I have necessary authorization to access all the data useful for my task.
    Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

Satisfaction

22. I am satisfied with the overall quality of information generated from the data warehouse.
    Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

23. I am satisfied with the overall performance of the data warehouse system in its functionality, flexibility, and processing speed.
    Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

    If you are not, does your dissatisfaction come from;
    Functionality [ ], or flexibility [ ], or processing speed [ ] ?

Ease of Use

24. The data warehouse system is convenient and easy to use.
    Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

User Training

25. I would like more training for me or my staff on how to find, understand, access or use the data warehouse system.
    Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

System Performance (Throughput)

26. I sometimes have trouble accessing the data warehouse and am forced to wait.
    Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

27. Once I access the data warehouse, I usually get the necessary information without much waiting.
    Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

Information Utility

28. The information from the data warehouse is useful and makes me more productive.
    Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

29. The information from the data warehouse is critical to my work.
    Strongly disagree 1[ ], 2[ ], 3[ ], 4[ ], 5[ ], 6[ ], 7[ ] Strongly agree

IV. Suggestions

30. Please describe all technical or non-technical problems encountered during the use of the data warehouse.

31. Please provide any suggestions related to the use of the current data warehouse system.

Appendix B
Summary of Confirmatory Factor Analysis

Constructs	Cronbach Alpha	Indicators	Factor Loading
Data Quality	0.822	14 15 16	.703 .727 .785
Ability to Locate Data	0.897	18 19 20	.756 .904 .827
Satisfaction	0.751	22 23	.803 .803
Performance	0.639	26R 27	.744 .744
Infomation Utility	0.726	28 29	.787 .787

Extraction method: Principal component analysis.
Rotations: varimax
"R" represents reversed data.

Appendix C
Spearman Rank Correlation Matrix

	Data Quality	Ability to Locate Data	Throughput
Data Quality	–	.552	.492
Ability to Locate Data		–	.473
Throughput			–

Journal of the Association for Information Systems

ISSN: 1536-9323