2008

A Structural Model of Data Warehousing Success

Hongjiang Xu
Butler University, hxu@butler.edu

Mark I. Hwang

A STRUCTURAL MODEL OF DATA WAREHOUSING SUCCESS

MARK I. HWANG
Central Michigan University
Mt. Pleasant, MI 48859

HONGJIANG XU
Butler University
Indianapolis, IN 46208

ABSTRACT

Data warehousing is an important area of practice and research, yet few studies have assessed its practices in general and critical success factors in particular. Although plenty of guidelines for implementation exist, few have been subjected to empirical testing. Furthermore, no model is available for comparing and evaluating the various claims made in different studies. In order to better understand the critical success factors and their effects on data warehousing success, a research model is developed in this paper. This model is useful for comparing findings across studies and for selecting variables for future research. The model is tested using data collected from a cross sectional survey of data warehousing professionals. Partial Least Square (PLS) is used to validate the structural relations identified in the model. The resulting model has three groups of success factors. Technical factor is found to positively influence information quality, whereas both operational and economic factors have a positive effect on system quality. Structural relations are also found among the dependent variables. System quality positively influences information quality, which in turn positively affects individual benefits. Individual benefits in turn has a positive relation with organizational benefits.

Keywords: Data warehousing, critical success factors, information systems success, modeling

INTRODUCTION

A data warehouse has been defined as "a subject-oriented, integrated, nonvolatile, and time-variant collection of data in support of management's decisions" (Inmon 2002, p. 31). It is developed by extracting data from various source systems, cleaning and transforming the data, and loading it in the warehouse where it is then made available to decision makers (Watson et al., 2001). Since the early 1990s, the data warehouse has become the foundation of advanced decision support applications (Shim et al., 2002). Using sophisticated online analytical processing (OLAP) and data mining tools, some corporations are able to exploit insights gained from their data warehouse to significantly increase sales (Cooper et al., 2000; Heun, 2000; Whiting, 1999), reduce costs (Watson and Haley, 1998; Whiting, 1999), and offer new and better products or services (Cooper et al., 2000; Levinson, 2000; Watson and Haley, 1998). The payoff from a well-managed data warehouse can be huge. For instance, a study conducted by IDC, a leading research firm, found the average return on investments in data warehousing projects to be about 400 percent (Desai, 1999). By the late 1990s, most large corporations had either built or were planning to build a data warehouse (Joshi and Curtis, 1999).

However, the implementation of a data warehouse is both very expensive and highly risky. Building and marinating a data warehouse routinely cost a corporation millions of dollars (Gagnon, 1999; Jukic, 2006). At the same time, success seems to be the exception rather than the rule (Dagan, 2007). One early study reported that one-half to two-thirds of all initial data warehousing efforts fail (Kelly, 1997), while another study placed the failure rate at 60 to 90 percent (Voelker, 2001). Despite the advancement in technology and experience in implementation over the last decade, it is not unusual to hear failure rates around 50 to 75 percent cited by practitioners and consultants (Beal, 2005; Madsen, 2005; Watson, 2005). The reason, according to Madsen (2005), is that "people keep making the same mistakes". Nevertheless, spending on data warehousing grew at a healthy 43 percent annually though 2003 (Trowbridge, 2000) and is expected to rise significantly in 2005 (Agosta, 2004). A major reason is that, with the dramatic drop in storage costs, companies are racing to build ever-larger data warehouses in pursuit of greater granularity and real time information. For instance, Harrah's Entertainment, a leader in data warehousing, is reportedly spending $10 million to build a 30-terabyte data warehouse (Lyons, 2004). Without a good grasp of the core data warehousing success issues, however, spending more money can potentially create bigger problems and result in expensive failures.

Like every major information systems (IS) project, any number of things can go wrong with a data warehousing endeavor. Unfortunately, the precise nature of the critical success factors and their impact on implementation is unclear (Mukherjee and D'Souza, 2003). Currently, there is no theoretically sounded model available for comparing and evaluating the various claims made in different studies. In order to fill this gap in the literature and better understand the critical success factors and their effects on data warehousing success, the purpose of this study is to first develop a theoretical model, and then test and modify the proposed model using empirical data. The final empirically validated research model should be of interest and useful to both data warehousing practitioners and researchers.

PRIOR STUDIES

Despite the recognition of data warehousing as an important area of practice and research, relatively few studies have been conducted to assess data warehousing practices in general and critical success factors in particular (Shin, 2003; Watson et al., 2001; Wixom and Watson, 2001). The literature is full of practitioners' accounts of data warehousing projects that have succeeded or failed and the possible reasons for these outcomes. Some attempts have been made to summarize their claims (e.g., Sakaguchi and Frolick, 1997; Vatanasombut and Gray, 1999); however, no generally accepted framework or model of data warehousing success exists. A few case studies have also investigated data warehousing implementation at selected companies (e.g., Cooper et al., 2000; Winter and Meyer, 2001; Watson, Fuller and Ariyachandra, 2004). It would be useful to test these findings in a cross-sectional survey (Watson, Fuller, and Ariyachandra, 2004). Results from all prior data warehousing success studies were used in this study to develop the research model discussed in the next section. The handful of data warehousing surveys that have been published to date are briefly reviewed in the next paragraphs.

Table 1 summarizes the five published survey studies, which differ widely in the variables measured. Some studies measured critical success factors while others measured data warehousing success; however, only one (Wixom and Watson, 2001) measured both critical success factors *and* data warehousing success. Without including both in the same study, the effect of any success factor on data warehousing success cannot be substantiated. Researchers have also defined and measured different success factors and data warehousing success variables. For example, user satisfaction was used as a measure for success in two studies (Chen et al., 2000; Shin, 2003), but not in the others (Watson et al., 2001; Wixom and Watson, 2001). The two studies conducted by Watson and colleagues used different success measures too. It appears that even the fundamental question of what constitutes data warehousing success has not been resolved.

Similar variations also exist among studies that measured success factors. When different factors are examined in different studies, how are the results to be compared across studies? As the reported result column in Table 1 shows, it is very difficult to compare research findings to pin down the exact critical success factors and their impact on data warehousing success. Our research model for conceptualizing the critical success factors *and* their effects on data warehousing success is presented next.

RESEARCH MODEL

Information systems success in general and data warehousing success in particular are multifaceted concepts. Similarly, scores of factors have been examined in various studies for their possible impact on data warehousing success. The research model developed for this paper is shown in Figure 1, which allows for conceptualization of data warehousing success, factors that contribute to the success, and the relationship between the two.

TABLE 1: Prior Data Warehousing (DW) Survey Studies

Study	DW Success Factors Measured	DW Success Measured	Results Reported	N
Watson & Haley (1997)	Upper management support, User, involvement, Having a business need, User support, Using a methodology, modeling, Defined, understandable goals, Good, clean data, Managing expectations	n.a.	Ordered list of success factors	121
Chen et al. (2000)	n.a.	Support for end users, Accuracy, format, and Preciseness, Fulfillment of end users needs, User satisfaction	Support for end users affects user satisfaction	42
Wixom & Watson (2001)	Management support, Champion, Resources, User participation, Team skills, Source systems, Development technology, Organizational implementation success, Project implementation success, Technical implementation success	System quality, Data quality, Net benefits	Some success factors affect DW success	111
Watson et al. (2001)	n.a.	Reduced effort by developers to produce information, Improved user ability to produce information, More and better information, Better decisions, Improvement for business process, Support for the accomplishment of strategic business objectives	Ordered list of success measures	106
Shin (2003)	n.a.	System quality, information quality, service quality, user satisfaction	System quality affects user satisfaction	64

The model posits that data warehousing success is represented by four dependent variables: system quality, information quality, individual benefits, and organizational benefits. The success is affected by four independent variables or success factors: operational, technical, schedule, and economic. This model is general and useful for comparing all the variables that have been discussed in the data warehousing literature. For example, the four success variables can be used to classify all data warehousing benefits discussed in prior studies. On the other hand, most prior studies have concentrated on success factors related to either operational or technical factors. However, given the enormous time and financial resources involved, more attention is paid in this model to the schedule and economic factors. Table 2 lists the eight variables and the 19 measures developed for testing the research model. Justification for the measures, variables, and the relations hypothesized is discussed next.

DATA WAREHOUSING SUCCESS

Various authors have touted numerous benefits or advantages of data warehousing since the early days of its development. In an early review of the literature, Sakaguchi and Frolick (1997) categorized the advantages of data warehousing cited in 456 articles into 16 types of benefits. In another review, Vatanasombut and Gray (1999) listed 12 goals of data warehousing that can be classified as financial, operational, or application. Watson and colleagues (Watson and Haley, 1997; Watson et al., 2001) conducted a series of data warehousing studies and developed a taxonomy of data warehousing benefits that classifies six benefits based on their ease of measurement and level of impact. Wixom and Watson (2001) investigated the effects of several implementation factors on three success variables. How does one reconcile all these different success measures?

The DeLone and McLean information systems success model (1992) has emerged as a dominant model in the selection of dependent variables by MIS researchers. This model classifies all success measures into six categories — system quality, information quality, use, user satisfaction, individual impact, and organizational impact. The model was later modified to include system quality, information quality, service quality, use, user satisfaction, and net benefits as success measures (DeLone and McLean, 2003).

DeLone and McLean (1992) recommend that researchers select proper measures from the success model based on their research context. So, the question is: Which variables are appropriate for data warehousing? Figure 1 shows that the model of this research includes four success variables — system quality, information quality, individual benefits, and organizational benefits — that encompass all the benefits or advantages mentioned previously (Sakaguchi and Frolick, 1997; Vatanasombut and Gray, 1999; Watson and Haley, 1997; Watson et al., 2001; Wixom and Watson, 2001). Use is excluded from the model because its utility in general (Seddon, 1997) and in data warehousing studies in particular (Wixom and Watson, 2001) has been questioned. User satisfaction is similarly excluded because it is not considered a good indicator of success for multiple-user applications such as data warehouses (Wixom and Watson, 2001). In addition, it's unlikely that companies would spend millions of dollars on data warehousing just to make their users "happy." Service quality is not selected because it is more suited for the evaluation of an IS function rather than a product such as a data warehouse (DeLone and McLean, 2003). Finally, in the modified IS success model, the individual and organizational impacts are combined into a net benefits variable. DeLone and McLean (2003) acknowledge that the performance impact of IS needs to be assessed at different levels in individual research studies; therefore, individual and organizational impacts are retained and renamed as individual and organizational benefits in Figure 1.

System Quality

In accordance with its focus on decision support, a successful data warehouse is generally characterized as easy to use and efficient in producing information useful to decision makers. Although some attractive features that apply to other systems, such as scalability, standardization, and security have been mentioned

TABLE 2: Research Variables and Measures

Variable	Measure 1	Measure 2	Measure 3	Measure 4
Operational factor	clearly defined business needs/benefits	top management support	user involvement/ participation	
Technical factor	source data quality	proper development technology	adequate IS staff and consultant	project management (teamwork)
Schedule factor	practical implementation schedule	proper planning / scoping of project		
Economic factor	adequate funding	measurable business benefits		
System quality	easy to use	speedy information retrieval		
Information quality	better quality information	improved productivity		
Individual benefits	improved productivity	better decisions		
Organizational benefits	improved business processes	increased competitive position		

(Sakaguchi and Frolick, 1997), the success of a data warehouse is more than likely be judged by how easy and efficient it is for both end users and IS professionals to generate information to support decision making (Nelson, Todd, and Wixom, 2005; Shin, 2003; Vatanasombut and Gray, 1999; Watson and Haley, 1997). On the other hand, a data warehouse that is not user-friendly in either its user interface or the analysis tools provided can result in millions of dollars of unused software and unrealized returns on investment (Gorla, 2003; Johnson, 2004). Consequently, the current research chose "easy to use" and "speedy information retrieval" to measure system quality.

Information Quality

Quality information is an important asset for organizations (Wang, Storey and Firth, 1995). It seems that a data warehouse is expected to enable production of information of higher quality as well as new information that may be put to innovative use. Sakaguchi and Frolick (1997), for instance, discuss one of the advantages of a data warehouse as its ability to provide quantitative values, or metrics that allow a company to benchmark performance in an effort to measure progress. In other words, both the quality and quantity of information are important. As described by Watson and Haley (1997), more and better information is one of the purported benefits of data warehousing. The quality or usefulness of information is also used by both Shin (2003) and Wixom and Watson (2001) as one of their success measures. Consequently, the current research chose "more information" and "better quality information" to measure information quality.

Individual Benefits

Researchers generally agree that proper use of a data warehouse can make its users more efficient and effective. Armed with more and better information, employees should be able to improve productivity and make better decisions (Sakaguchi and Frolick, 1997; Vatanasombut and Gray, 1999; Watson and Haley, 1997; Watson, Fuller, and Ariyachandra, 2004). Consequently, the current research chose "improved productivity" and "better decision" to measure individual benefits.

Organizational Benefits

This is where data warehousing is purported to offer the greatest payoff. Proper use of a data warehouse can help achieve all sorts of strategic advantages by lowering costs, increasing revenues, improving business processes, and supporting initiatives such as customer relationship management and knowledge management (Sakaguchi and Frolick, 1997; Vatanasombut and Gray, 1999; Watson and Haley, 1997; Watson, Fuller, and Ariyachandra, 2004). Consequently, the current research chose "improved business processes" and "increased competitive position" to measure organizational benefits.

As shown in Figure 1, four positive relations are hypothesized among the four success variables. System quality is expected to positively affect individual benefits, based on prior studies in both data warehousing (Wixom and Watson, 2001) and other systems (Etezadi-Amoli and Farhoomand, 1996; Goodhue and Thompson, 1995; Teo and Wong, 1998). The same relation was

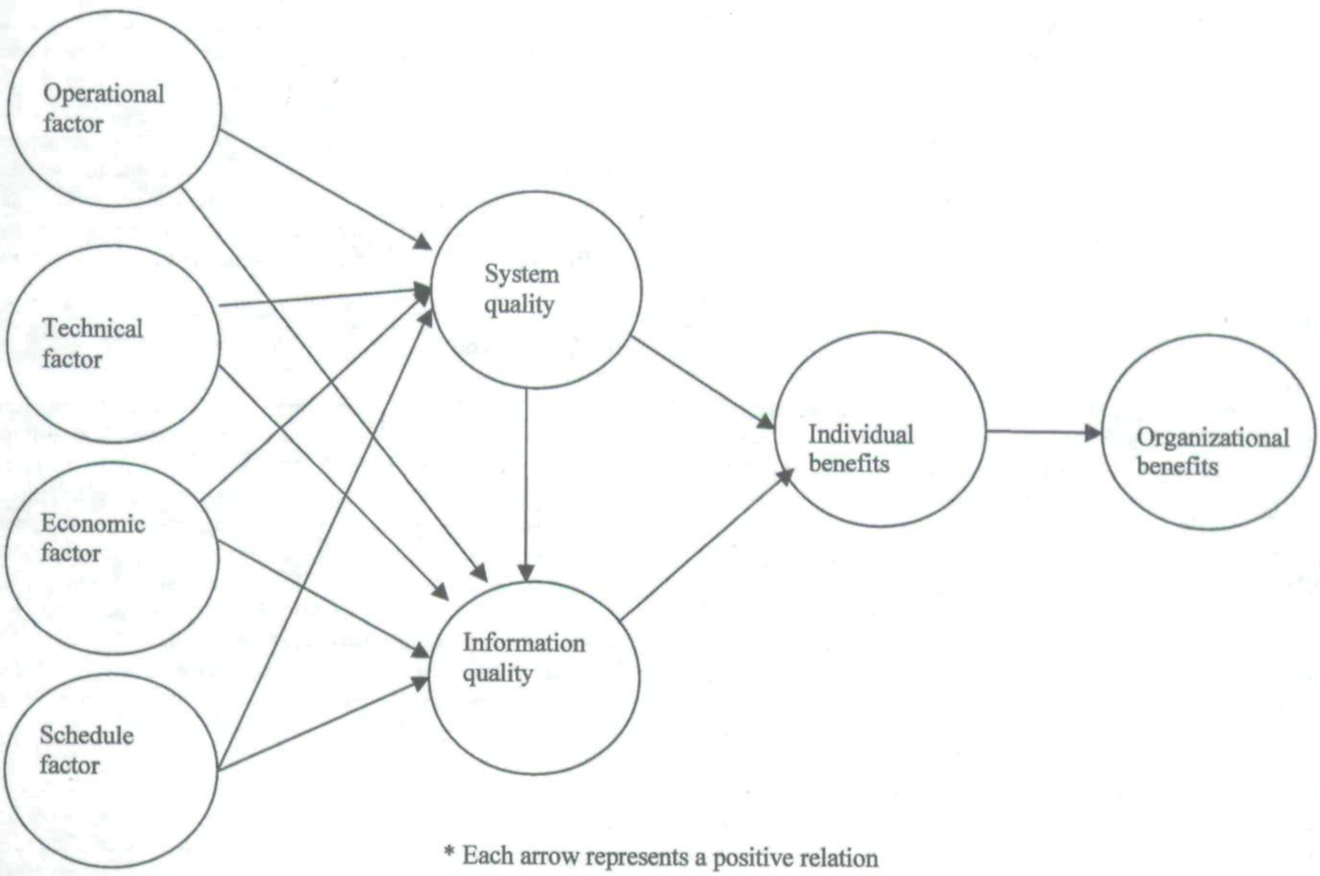


FIGURE 1. Research Model

also found between information quality and individual benefits in prior studies (Etezadi-Amoli and Farhoomand, 1996; Teo and Wong, 1998; Wixom and Watson, 2001). A positive relationship between individual benefits and organizational benefits was similarly supported in prior IS studies (Jurison, 1996; Teo and Wong, 1998). To the extent that system quality represents the desired characteristics of an information system and that information quality represents the desired characteristics of the product of an information system (DeLone and McLean, 1992), it is conceivable that system quality has a positive impact on information quality. Interestingly, the relation between system quality and information quality is *not* postulated in the DeLone and McLean model; nor has it been studied in the IS success literature. However, in data quality research, it has been established that data or information quality and system quality are related. For instance, Wang and Strong (1996) propose that data quality has four distinct dimensions: intrinsic, contextual, representational, and accessibility. Intrinsic quality denotes that data has value in its own right whereas contextual value comes from its use in the context of the task at hand. Representational and accessibility are related to how information is presented and accessed, both of which "emphasize the importance of the role of systems" (p. 6). Similarly, Orr (1998) considers data quality fundamentally intertwined in how a system fits into the real world. Even though system and information quality may be conceptualized or measured differently, a positive relation is plausible and thus hypothesized in this research.

CRITICAL SUCCESS FACTORS

There has long been a keen interest in identifying the factors that contribute to the success or failure of data warehouses. In an early review of the literature, Vatanasombut and Gray (1999) identified 51 success factors that may be classified into 12 categories. Most of these 51 factors, however, apply not only to data warehousing, but also to large systems development projects in general; only nine factors are specific to data warehousing. Other researchers provided their own lists of critical success factors. For example, Watson and Haley (1997) identified eight critical success factors, whereas Sammon and Finnegan (2000) discussed their "ten commandments of data warehousing." Although the identification of various factors for data warehousing success is helpful, it again raises the question of how to reconcile all these different critical success factors.

Using the four feasibility tests applied to traditional systems development projects, factors that affect data warehousing outcomes can be classified as operational, technical, schedule, or economic factors. Most factors discussed in the data warehousing literature fall into either the operational or technical categories. The former includes top management support (Gardner, 1998; Sammon and Finnegan, 2000; Vatanasombut and Gray, 1999; Watson and Haley, 1997; Wixom and Watson, 2001; Watson, Fuller, and Ariyachandra, 2004) and a business driver (Baker and Baker, 1999; Sammon and Finnegan, 2000; Watson, Fuller and Ariyachandra, 2004), whereas the latter includes data quality (Beal, 2005; Joshi and Curtis, 1999; Sammon and Finnegan, 2000; Wixom and Watson, 2001; Watson, Fuller, and Ariyachandra, 2004). However, as with any large-scale IS project, time and money are critical issues that need to be dealt with properly. For instance, there are increasing calls to assess data warehouse ROI in the literature (Lewis, 2001; Sinn, 2003; Whiting, 1999).

Operational Factor

This factor measures how well the system solution fits the problem. It is also concerned with the role played by management and users during implementation, as well as with their perceptions of the new system. Many researchers have stressed the importance of having a business driver for a data warehouse (e.g., Baker and Baker, 1999; Sammon and Finnegan, 2000). Top management support is critical to all major IS initiatives and has been noted for its importance in data warehouse development as well (Watson and Haley, 1997; Wixom and Watson, 2001). User involvement/participation is important to IS projects in general (Hwang and Thorn, 1999) and data warehousing in particular (Conner, 2003; Watson and Haley, 1997; Wixom and Watson, 2001). Consequently, the current research chose "clearly defined business needs/benefits," "top management support," and "user involvement/participation" to measure the operational factor.

Technical Factor

This factor measures the availability of technical resources and expertise. It is concerned with both the maturity of the technology and the availability of technical expertise in-house. Many companies choose to utilize consultants or third party vendors for their data warehousing projects due to technical considerations. One of the success factors cited by Cooper et al. (2000) in the implementation of a data warehouse at a major bank was the replacement of the in-house development team with outside consultants. Many companies have also brought in outside consultants to bring a stalled project back on track (Connor, 2003). Almost all authors emphasize the technical aspects of data warehousing projects, including cleansed data, meta data, standard methodology, and project management as very critical to the success of the project (e.g., Baker and Baker, 1999; Joshi and Curtis, 1999; Sammon and Finnegan, 2000; Vatanasombut and Gray, 1999; Watson and Haley, 1997; Wixom and Watson, 2001). Consequently, the current research chose "source data quality," "proper development technology," "adequate IS staff and consultants," and "project management/teamwork" to measure the technical factor.

Schedule Factor

The schedule factor measures how reasonable the time allowed for development of an information system is. Deadlines may be mandatory or desirable. Mandatory deadlines are usually the result of new laws or regulations and thus do not apply to data warehousing projects. Nevertheless, proper planning and execution of the implementation schedule may be critical to data warehousing success (Baker and Baker, 1999; Sigal, 1998; Watson, Fuller, and Ariyachandra, 2004). Moreover, "scope creep" is a common cause of project failure (Conner, 2003). Consequently, the current research chose "practical implementation schedule" and "proper planning/scoping of project" to measure the schedule factor

Economic Factor

The economic factor, also known as cost-benefit analysis, measures the bottom line. This type of analysis is usually performed for transaction processing system projects that can easily quantify benefits. Data warehouses are mostly created for decision support or strategic applications that do not have apparent measurable benefits. Consequently, the economic factor was not a priority in early projects. However, as the technology matures and experience is gained, more and more companies are conducting some type of cost-benefit analysis (Lewis, 2001; Sinn, 2003; Whiting, 1999). As a result, the current research chose "adequate funding" and "measurable business benefits". to measure the economic factor.

As shown in Figure 1, eight positive relations are hypothesized among the success factors and system and information quality. Each factor is expected to contribute to the development of a "successful" data warehouse, as discussed above. The specific effect of any success factor on any success variable; however, has not been empirically tested. The only exception that we know of is Wixom and Watson (2001), which found a set of "implementation factors" positively affect both system quality and information quality. To the extent that system quality represents the desired characteristics of a system and that information quality represents the desired characteristics of the product of the system (DeLone and McLean, 1992), all factors are expected to positively affect both system quality and information quality.

METHODOLOGY

Based on the research model discussed above, a web-based questionnaire was developed to collect data on the 19 measures perceived by data warehousing professionals. Surveys are a common approach for collecting the large amounts of data needed for statistical testing of relationships. They have been employed to study critical success factors in other information systems research areas, such as information centers (Magal, Carr and Watson, 1988). This approach is not without its limitations; however, such as the possible interaction of the factors (Nandhakumar, 1995) and the exclusion of context variables including social, cultural, political, and economic factors. (Bussen and Myers, 1997). These caveats should be kept in mind when interpreting survey results.

The Data Warehousing Institute E-mail List was used as the source of the survey. This list contains the contact information of over 15,000 data warehousing professionals. An e-mail was sent to a random sample of 6,000 recipients. A $10 gift certificate from Amazon.com was used as an incentive for participation. A follow-up email was sent three weeks later. The two rounds of mailings yielded 98 completed questionnaires. The 1.6 percent response rate obtained in this study is typical of unsolicited mailings sent out by our e-mail list vendor, whose response rates range from one to two percent. The total number of respondents is at the high end of sample sizes reported in prior surveys (see Table 1).

The respondents were asked to rate how significant each of the eight success measures was using a five-point scale. A sample question is "How significant is easy to use to your warehouse?" They were also asked to rate how important each of the eleven success factor measures was using a five-point scale. A sample question is "How important is clearly defined business needs/ benefits to the success of your warehouse?" The responses were analyzed using partial least square (PLS), a structural modeling technique.

PLS was chosen over other structural modeling approaches such as LISREL because of its ability to handle formative measures. All the measures used in this research are formative; i.e., they cause rather than are caused by the underlying constructs. In addition, PLS allows for the testing of the psychometric properties of the measures (the measurement model) and the relations among the variables (the structural model) simultaneously. Testing of the measurement model enabled us to adjust some of the measures in the validation of our research model, as explained later. PLS Graph version 3.0 (Chin, 2001) was used for analysis, and the bootstrap resampling method (100 resamples) was used to test the significance of the structural relations.

RESULTS

Demographics

The largest group of respondents (38%) was data warehousing specialists, followed by DBA's (21%). The remaining respondents were managers (18%), consultants (16%), and analysts (6%). They worked in a variety of industries with the largest category being consulting/professional services (16%), followed by federal government (11%). The largest group of these organizations (33%) had annual revenue less than 10 million dollars, while the next largest portion (18%) had annual revenue between 100 and 500 million. Over one third of these data warehouses (31%) took from six to 12 months to develop, whereas the next largest group (25%) took from 12 to 24 months. The largest group of these data warehouses (22%) was deployed two years ago, and the next largest category (20%) was deployed three years ago. Finally, over one third of these data warehouses (32%) were less than 100 gigabytes. The next common categories were from 100 to 500 gigabytes and from 500 gigabytes to one terabyte, both accounting for 16 percent.

Measurement Model

The desired psychometric properties of a model include internal consistency, convergent and discriminant validity. Internal consistency is demonstrated when the reliability of each measure is above 0.70 (Nunally, 1978). This is true of all the measures taken in the survey. Convergent validity is adequate when each construct has an average variance extracted (AVE) of at least 0.5 (Fornell and Larcker, 1981). Two of the constructs, operational and technical factors, did not satisfy this condition when the research model was first tested, indicating possible measurement errors. In addition, some of the path coefficients were not significant, e.g., the path from the schedule factor to system quality and that from the schedule factor to information quality. While it is not uncommon to find non-significant results in any research project, having a construct not related to others in a structural model is not meaningful. To improve the model, therefore, the most non-significant measures were dropped, which included top management support and proper planning/scoping. The remaining schedule measure, practical implementation schedule, was reassigned to the economic factor. To the extent that time is money, it makes sense to consider implementation schedule part of the economic factor. The resultant model is shown in Figure 2. All the constructs have an AVE of over 0.5.

Discriminant validity is acceptable if the AVE of each construct is greater than the variance among all constructs (Chin, 1998). This is usually demonstrated by showing that the square root of an AVE is greater than the correlations among the construct and all other constructs in the model. All the constructs in the revised model have satisfactory discriminant validity.

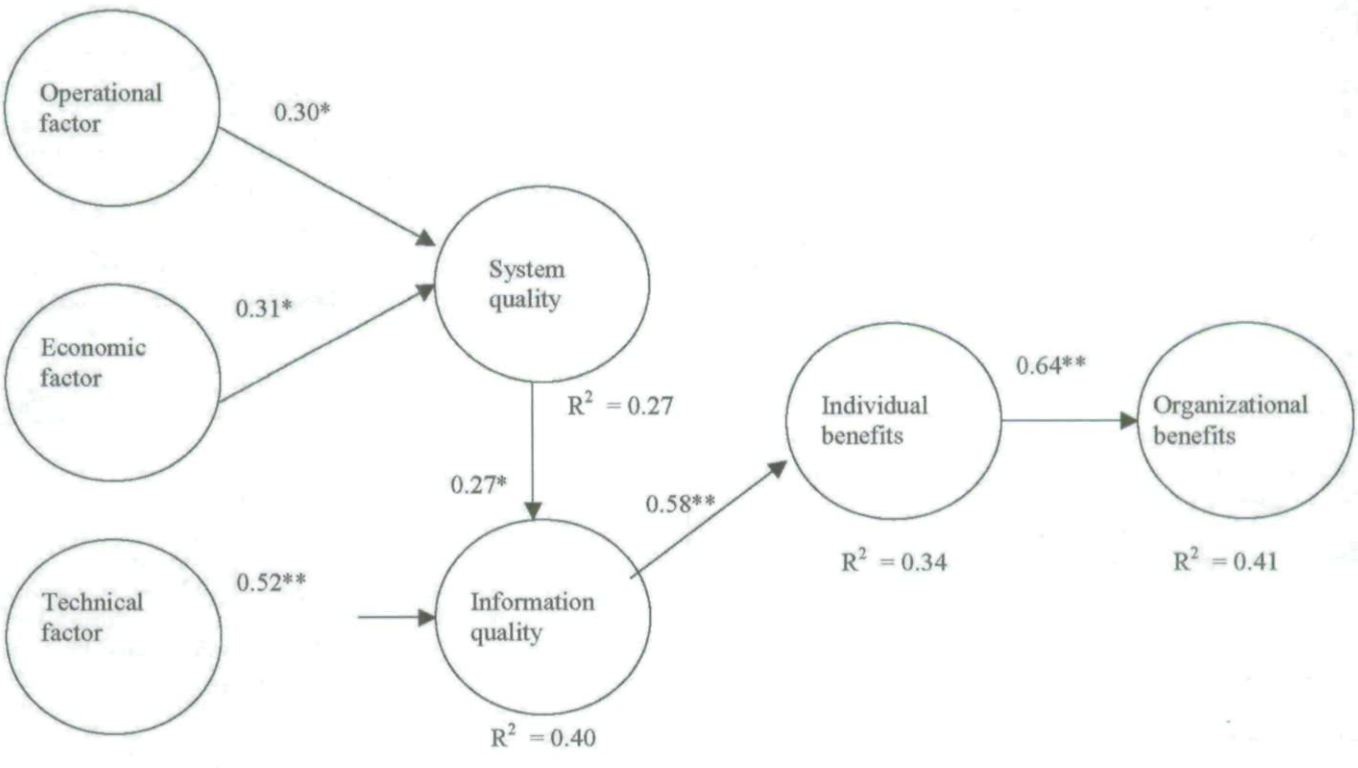
Structural Model

Figure 2 shows the estimated path coefficients and R^2 values, which together describe the structure of the model. The path coefficients and R^2 are interpreted the same way as are regression coefficients and variance explained in regular regression models. As expected, all the path coefficients are positive and significant at either the five percent or the one percent level. Operational (0.30) and economic factors (0.31) had about equal influence on system quality. The effect of technical factor, on the other hand, was manifested in information quality. System quality in turn had a positive effect on information quality. Information quality had a positive effect on individual benefits, which in turn had a positive effect on organizational benefits.

DISCUSSION

This research contributes to the understanding of data warehousing success by showing the interrelationships among a set of variables. Building on the IS success model of DeLone and McLean (1992) and data quality research, this study has found that the quality of a data warehouse has a positive effect on the quality of its product — information. The consumption of quality information, in turn, results in benefits to its recipients. These individual benefits, in turn, leads to positive organizational outcomes or benefits. Various advantages of data warehouses have been touted in the literature for some time, but this is a rare piece of evidence supporting their existence *and* their relationships. These advantages are related in a hierarchical fashion with system quality being the most fundamental. This is important because it points out that higher order payoffs cannot be expected until more basic benefits are realized. It is, therefore, prudent to ignore the dramatic productivity claims made by software vendors and instead concentrate on more fundamental system and information quality issues. In a study of information and system quality, Nelson, Todd, and Wixom (2005) found accuracy the most important aspect of information quality. They also caution against the assumption that data warehousing universally produces high-quality information. The importance of data quality is similarly stressed by Gartner Group, which cites poor data quality as the main cause of high data warehouse project failure rates (Beal, 2005). Likewise, no firm-level benefits ought to be expected until an organization's employees have reaped the benefits individually from the use of a data warehouse.

This research has also shown the specific effects of factors critical to data warehousing success. Even though prior studies have discussed numerous critical success factors, this is a rare piece of evidence showing how success factors are related to data warehousing success. Both operational and economic factors have a positive effect on system quality. Wixom and Watson (2001) found a similar, though indirect, effect of comparable factors on system quality. This makes sense since these factors are supposed to facilitate the delivery of a quality data warehouse. The effect of the technical factor, however, is more salient on information quality than on system quality. This may have to do with the make up of the technical factor, whose most significant measure happens to be source data quality. Alternatively, it may mean that technical issues in data warehousing are related more to information than to the system itself. Interestingly, none of the factors examined by Wixom and Watson (2001) had a significant effect on information quality. This area warrants further investigation.



* $P < 0.05$; ** $P < 0.01$

FIGURE 2. Validated Data Warehousing Success Model

Two measures were dropped in the model validation process. Top management support, despite its prominence in the literature, may not be important to data warehousing after all. This at first seems surprising and may mean that top management has fully bought into the idea of data warehousing, and, therefore, is no longer a critical factor. An alternate explanation is that it is still important but its effect is indirect, and, thus, not significant in our model. This view is partially supported by Wixom and Watson (2001), who reported that top management support affects organizational implementation success, which in turn affects system quality. The indirect effect of top management support on data quality, however, was not supported (Wixom and Watson, 2001). Watson, Fuller, and Ariyachandra (2004) offer another perspective. In their case study of a data warehouse implementation at an insurance company, they found that mere participation by senior management was not sufficient for success; the management needed to be passionately involved. The second measured dropped was proper planning/scoping of project. This again may mean that the measure is unimportant or that alternate definition or operationalization is needed in future research.

CONCLUSIONS

Data warehousing success is an important issue for both researchers and practitioners; however, not many studies have empirically assessed data warehousing practices in general and critical success factors in particular. Although plenty of guidelines for implementation exist, few have been subjected to rigorous empirical testing. Another problem is that researchers have used different variables in individual studies, thus making comparison and integration of the results from different studies difficult. This paper develops a research model for data warehousing success to facilitate research integration and variable selection in future research. The model is general and new variables or measures, when identified, can be added easily. For example, as companies race to build ever-larger warehouses in pursuit of greater granularity and real time information, backing up terabytes of data can be a challenge. In such an environment, "easy to manage" could become a success measure, which can be an addition to the "system quality" variable of the model.

As mentioned earlier, most prior studies have examined either critical success factors or data warehousing success, but not both. Researchers are encouraged to start including both sets of variables to test the effect of any critical success factor. Relations that are not supported in this research (e.g., the path from technical factor to system quality) require further study; so are relations that are supported when new measures are developed (e.g., a specially designed query processor). We believe that the list of critical success factors could vary as the timeframe or the environment changes. In executive information systems (EIS) research, Nandhakumar (1996) argued that the success factors may interact and that their effects may vary during different stages of a project. Bussen and Myers (1997), in their study of an EIS implementation, similarly concluded that satisfying a static set of factors is not sufficient as an explanation for system outcomes. In data warehousing research, Doherty and Doig (2003) concluded in a case study that the success of a data warehouse implementation depends on how well the resulting culture changes are measured and managed. The potential impact of cultural changes, as well as other political, social, and economic factors (Bussen and Myers, 1997) should be further researched to allow a fuller understanding of data warehousing success. The ongoing study of data warehousing success is worthwhile, and researchers should find the research model discussed in this paper useful.

REFERENCES

Agosta, L. (November 2004) Data warehousing lessons learned: A time of growth for data warehousing. *DM Review Magazine*. Retrieved on 7/24/2007 at http://www.dmreview.com/article_sub.cfm?articleId=1012461.

Baker, S. and Baker, K. (1999). The best little warehouse in business. *Journal of Business Strategy*, 20, 2, 32-37.

Beal, B. (2005). Report: Half of data warehouse projects to fail. CRMNews. Retrieved on 7/24/07 at http://searchcrm.techtarget.com/originalContent/0,289142,sid11_gci1066086,00.html.

Bussen, W., and Myers, M.D. (1997). Executive information system failure: A New Zealand case study. *Journal of Information Technology*, 12, 145-153.

Chen, L., Soliman, K., Mao, E., and Frolick, M. (2000). Measuring user satisfaction with data warehouses: An exploratory study. *Information & Management*, 37, 103-110.

Chin, W. (1998). Issues and opinions on structural equation modeling. MIS Quarterly, 22, 1, vii-xvi.

Chin, W. (2001). *PLS-Graph User's Guide, Version 3.0*. Soft Modeling.

Conner, D. (2003). Data warehouse failures commonplace. *Network World*, 20, 3, 24.

Cooper, B.L., Watson, H.J., Wixom, B.H., and Goodhue, D.L. (2000). Data warehousing supports corporate strategy at First American Corporation. *MIS Quarterly*, 24, 4, 547-567.

Dagan, B. (2007). Why Has "Data Warehouse" Become a Dirty Word? *Natural Gas and Electricity*, 23, 12, 18-22.

DeLone, W. and McLean, E. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research*, 3, 1, 60-95.

DeLone, W. and McLean, E. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems*, 19, 4, 9-30.

Desai, A. (1999). For pharmaceutical companies, a data warehouse can be just what the doctor ordered. *Health Management Technology*, 20, 2, 20-22.

Doherty, N.F. and Doig, G. (2003). An analysis of the anticipated cultural impacts of the implementation of data warehouses. *IEEE Transactions on Engineering Management*, 50, 1, 78-88

Etezadi-Amoli, J., and Farhoomand, A.F. (1996). A structural model of end user computing satisfaction and user performance. *Information & Management*, 30, 2, 65-73.

Fornell, C., and Larcker, D.F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18, 3, 39-50.

Gagnon, G. (1999). Data warehousing: An overview. *PC Magazine*, March 19, 245-246.

Gardner, S. (1998). Building the data warehouse. *Communications of the ACM*, 41, 9, 52-60.

Goodhue, D.L., and Thompson, R.L. (1995). Task-technology fit and individual performance. *MIS Quarterly*, 19, 2, 213-233.

Gorla, N. (2003). Features to consider in a data warehousing system. *Communications of the ACM*, 46, 11, 111-115.

Heun, C.T. (2000). Harrah's bet on IT to understand its customers. *InformationWeek*, 816, December 11, 10-12.

Hwang, M.I., and Thorn, R. (1999). The effect of user engagement on system success: a meta-analytical integration of research findings. *Information & Management*, 35, 4, 229-236.

Inmon, W. (2002). *Building the Data Warehouse*, 3rd ed. New York: John Wiley & Sons.

Johnson, L.K. (2004). Strategies for data warehousing. *MIT Sloan Management Review*. 45, 3, 9.

Joshi, K. and Curtis, M. (1999). Issues in building a successful data warehouse. *Information Strategy*, 15, 2, 28-35.

Jukic, . (2006). Modeling strategies and alternatives for data warehousing projects. *Communications of the ACM*, 49, 4, 83-88.

Jurison, J. (1996). The temporal nature of IS benefits: A longitudinal study. *Information & Management*, 30, 2, 75-79.

Kelly, S. (1997). *Data Warehousing in Action*, John Wiley & Sons, Chichester.

Levinson, M. (2000). Slices of lives. *CIO*, 13, 21, 126-136.

Lewis, D. (2001). Data warehousing: Planning key to payoff. *InternetWeek*, 853, 52.

Lyons, D. (December 13, 2004). Too much information. Forbes, 110-115.

Madsen, M. (2005). A 50% data warehouse failure rate is nothing new. Bounded Rationality: Analysis of Architecture, Integration, BI & IT Strategy. Retrieved on 7/24/07 at http://blogs.ittoolbox.com/eai/rationality/archives/a-50-data-warehouse-failure-rate-is-nothing-new-4669

Magal, S. R., Carr, H. H., and Watson, H. J. (1988). Critical success factors for information center managers. *MIS Quarterly*, 12, 3, 413-425.

Mukherjee, D., and D'Souza D. (2003). Think phased implementation for successful data warehousing. *Information Systems Management*, 82-90.

Nandhakumar, J. (1996). Design for success? Critical success factors in executive information systems development. *European Journal of Information Systems,* 5, 62-72.

Nelson, R.R., Todd, P.A., and Wixom, B.H. (2005). Antecedents of information and system quality: An empirical examination within the context of data warehousing. *Journal of Management Information Systems*, 21, 4, 199-235.

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

Orr, K. (1998). Data quality and system theory. *Communication of the ACM*, 41, 2, 66-71.

Sammon, D., and Finnegan, P. (Fall 2000). The ten commandments of data warehousing. *Database for Advances in Information Systems*, 31, 4, 82-91.

Sakaguchi, T. and Frolick, M. (1997). A review of the data warehousing literature. *Journal of Data Warehousing*, 2, 1, 34-54.

Seddon, P. (1997). A respecification and extension of the DeLone and McLean model of IS success. *Information Systems Research*, 8, 3, 240-253.

Shim, J.P., Warkentin, M., Courtney, J.F., Power, D.J., Sharda, R., and Carlsson, C. (2002). Past, present, and future of decision support technology. *Decision Support Systems*, 33, 2, 111-126.

Shin, B. (2003). An exploratory investigation of system success factors in data warehousing. *Journal of the Association for Information Systems*, 4, 141-170.

Sigal, M. (1998). A common sense of development strategy. *Communications of the ACM*, 41, 9, 42-43.

Sinn, B. (2003). A hunt for treasure? *Best's Review*, 103, 10, 90.

Teo, T.S.H., and Wong, P.K. (1998). An empirical study of the performance impact of computerization in the retail industry. *Omega-The International Journal of Management Science*, 26, 5, 611-621.

Trembly, A. (2001). Experts: Technology is not to blame for data warehouse failures. *National Underwriter*, 105, 45, 32, 41.

Trowbridge, D. (2000) Database overload overwhelms database administrators. *Computer Technology Review*, 20, 6, 20, 22.

Vatanasombut B., and Gray, P. (1999). Factors for success in data warehousing: What the literature tells us. *Journal of Data Warehousing*, 4, 3, 25-33.

Voelker, M. (2001). Databases: The next generation. *Insurance & Technology*, 26, 3, 30-34.

Wang, R. Y., Storey, V.C., and Firth, C.P. (1995). A framework for analysis of data quality research. *IEEE Transactions on Knowledge and Data Engineering*, 7, 4, 623-640.

Wang, R.Y., and Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12, 4, 5-33.

Watson, H. (2005). Are Data Warehouses Prone to Failure? *Business Intelligence Journal*, 10, 4, 4-6.

Watson, H., Annino, D., Wixom, B., Avery, K. and Rutherford, M. (2001). Current practices in data warehousing. *Information Systems Management,* 18, 1, 47-55.

Watson, H.J., Fuller, C., & Ariyachandra, T. (2004). Data warehouse governance: Best practices at Blue Cross and Blue Shield of North Carolina. *Decision Support Systems,* 38, 435-450.

Watson, H., and Haley, B. (1997). Data warehousing: A framework and survey of current practices. *Journal of Data Warehousing*, 2, 1, 10-17.

Watson, H., and Haley, B. (1998). Managerial considerations. *Communications of the ACM*, 41, 9, 32-37.

Whiting, R. (1999). Warehouse ROI. *InformationWeek*, 735, May 24, 99-104.

Winter, R., and Meyer, M. (2001). Organization of data warehousing in large service organizations: A matrix approach based on data ownership and competence centers. *Journal of Data Warehousing*, 6, 4, 23-29.

Wixom, B.H. and Watson, H.J. (2001). An empirical investigation of the factors affecting data warehousing success. *MIS Quarterly*, 25, 1, 17-41.