

2007

The Effect of Implementation Factors on Data Warehousing Success: An Exploratory Study

Hongjiang Xu

Butler University, hxu@butler.edu

Mark I. Hwang

Follow this and additional works at: http://digitalcommons.butler.edu/cob_papers



Part of the [Management Information Systems Commons](#)

Recommended Citation

Xu, Hongjiang and Hwang, Mark I., "The Effect of Implementation Factors on Data Warehousing Success: An Exploratory Study" (2007). *Scholarship and Professional Work - Business*. Paper 77.

http://digitalcommons.butler.edu/cob_papers/77

This Article is brought to you for free and open access by the College of Business at Digital Commons @ Butler University. It has been accepted for inclusion in Scholarship and Professional Work - Business by an authorized administrator of Digital Commons @ Butler University. For more information, please contact fgaede@butler.edu.

The Effect of Implementation Factors on Data Warehousing Success: An Exploratory Study

Mark I. Hwang and Hongjiang Xu
Central Michigan University, Mount Pleasant, MI, USA

mark.hwang@cmich.edu hongjiang.xu@cmich.edu

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. In order to better understand implementation factors and their effect on data warehousing success, perceptions of data warehousing professionals are examined in a cross sectional survey. Best subsets regression is used to identify the specific factors that are important to each success variable. Since different companies may have different objectives or emphases in their data warehousing endeavors, the results are useful in identifying the exact factors that need attention and in providing a basis for prioritizing those factors. The results also suggest several promising directions for continued research on data warehousing success.

Keywords: Data warehousing, critical success factors, IS implementation, system success, survey

Introduction

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, Watson, Wixom, & Goodhue, 2000; Heun, 2000; Whiting, 1999), reduce costs (Watson & Haley, 1998; Whiting, 1999), and offer new and better products or services (Cooper et al., 2000; Levinson, 2000; Watson & 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 & Curtis, 1999).

However, the implementation of a data warehouse is both very expensive and highly risky. One study reported an average cost of \$2.2 million for a typical data warehouse (Gagnon, 1999). At

Material published as part of this publication, either on-line or in print, is copyrighted by the Informing Science Institute. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact Publisher@InformingScience.org to request redistribution permission.

the same time, success seems to be the exception rather than the rule. 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). A more recent study put the failure rate around 41 percent (Connor, 2003). Despite the high failure rates, 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 all major information systems (IS) projects, any number of things can go wrong in a data warehousing endeavor. Unfortunately, the precise nature of the success factors and their impact on data warehousing are still unclear (Mukherjee & D'Souza, 2003). While many implementation factors that could contribute to success or failure have been discussed by practitioners and researchers alike, the effect of these factors has rarely been tested in empirical research. In order to fill this gap in the literature, the current study identifies a number of implementation factors and tests their effect on success using data collected from a cross sectional survey of data warehousing professionals. The objective is to produce an empirically validated list of factors and report their respective strength of impact on various aspects of data warehousing success. This list is readily useful to practitioners in their planning and implementation of data warehousing projects. The list also points out promising directions for continued research into data warehousing success.

Literature Review

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, Annino, Wixom, Avery, & Rutherford, 2001; Wixom & 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 & Frolick, 1997; Vatanasombut & Gray, 1999). A few case studies have investigated data warehousing implementation at selected companies (e.g., Cooper et al., 2000; Winter & Meyer, 2001; Watson, Fuller & Ariyachandra, 2004). It would be useful to test the claims made by practitioners and findings from case studies in a cross-sectional survey (Watson, Fuller, & Ariyachandra, 2004). The handful of data warehousing surveys that have been published to date are briefly reviewed in the next paragraphs.

Table 1 summarizes the six published survey studies, which differ widely in the variables measured. Some studies measured implementation factors while others measured data warehousing success; however, only one (Wixom & Watson, 2001) measured both implementation factors and data warehousing success. Without including both in the same study, the effect of any factor on data warehousing success cannot be substantiated. Researchers have also defined and measured different implantation factors and data warehousing success variables. For example, user satisfaction was used as a variable for success in two studies (Chen et al., 2000; Shin, 2003), but not in the others (Watson et al., 2001; Wixom & Watson, 2001). The two studies conducted by Watson and colleagues used different success variables 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 implementation 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 implementation factors and their impact on data warehousing success.

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	Not Available (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 info; Improved user ability to produce info; More and better info; Better decisions; Improvement for business process; Support for the accomplishment of strategic business objectives	Ordered list of success measures	106
Hwang & Cappel (2002)	N/A	N/A	Development/management Practices	27
Shin (2003)	N/A	System quality; Information quality; Service quality; User satisfaction	System quality affects user satisfaction	64

The Research Model

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 & 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. We selected eight success variables by consolidating the lists from prior data warehousing studies, as shown in Figure 1.

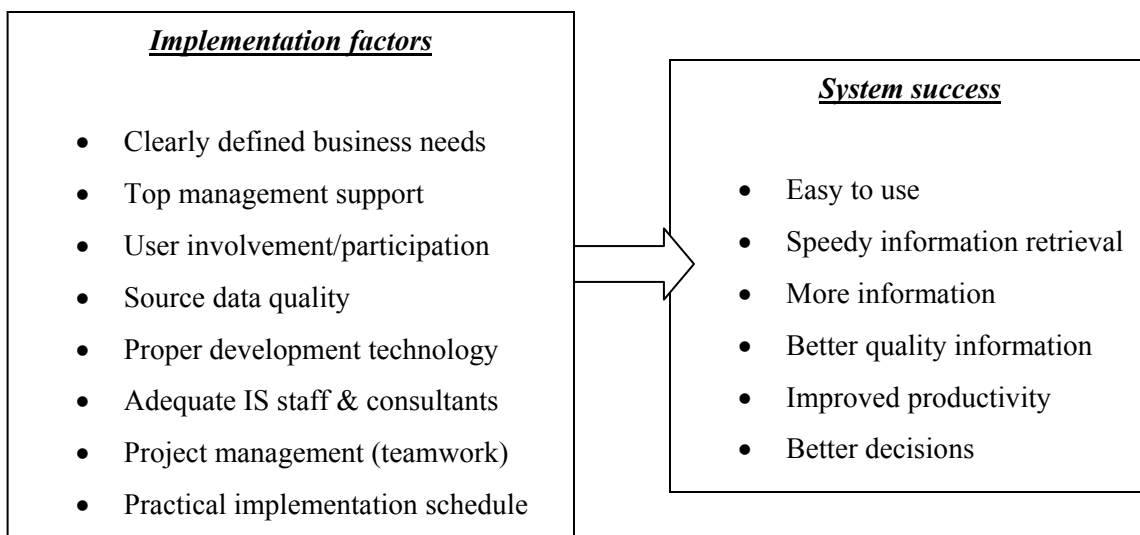


Figure 1: Research Model for Data Warehousing Success

System Success

The first two variables “easy to use” and “speedy information retrieval” measure the quality of the system developed. 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 (Sakaguchi & 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 (Shin, 2003; Vatanasombut & Gray, 1999; Watson & 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).

The next two variables “more information” and “better quality information” measure the benefits reflected in the output of a system--information. 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 variables.

The next two variables “improved productivity” and “better decision” measure benefits resulted from the use of information by individual decision makers. 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 & Frolick, 1997; Vatanasombut & Gray, 1999; Watson, Fuller, & Ariyachandra, 2004; Watson & Haley, 1997).

The final two variables “improved business processes” and “increased competitive position” measure benefits accrued at the organizational level. A well run data warehouse is purported to offer the greatest payoff at the organizational level by lowering costs, increasing revenues, improving business processes, and supporting initiatives such as customer relationship management and knowledge management (Sakaguchi & Frolick, 1997; Vatanasombut & Gray, 1999; Watson, Fuller, & Ariyachandra, 2004; Watson & Haley, 1997).

Implementation 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.”

We selected eleven factors based on these prior lists and a review of other data warehousing studies.

The first three factors “clearly defined business needs/benefits,” “top management support,” and “user involvement/participation” measure the operational aspect of the project. Many researchers have stressed the importance of having a business driver for a data warehouse (e.g., Baker & Baker, 1999; Sammon & 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 & Haley, 1997; Wixom & Watson, 2001). User involvement/participation is important to IS projects in general (Hwang & Thorn, 1999) and data warehousing in particular (Conner, 2003; Watson & Haley, 1997; Wixom & Watson, 2001).

The next four factors “source data quality,” “proper development technology,” “adequate IS staff and consultants,” and “project management/teamwork” measure the availability of technical resources and expertise for the project. 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 & Baker, 1999; Joshi & Curtis, 1999; Sammon & Finnegan, 2000; Vatanasombut & Gray, 1999; Watson & Haley, 1997; Wixom & Watson, 2001).

The next two factors “practical implementation schedule” and “proper planning/scoping of project” measure how reasonable the time allowed for development of a data warehouse is. Proper planning and execution of the implementation schedule is critical to data warehousing success (Baker & Baker, 1999; Sigal, 1998; Watson, Fuller, & Ariyachandra, 2004). Moreover, “scope creep” has been cited as a common cause of data warehousing project failure (Conner, 2003).

The last two factors “adequate funding” and “measurable business benefits” measure the economic aspect of the project. The economic factor, also known as cost-benefit or ROI 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; Watson et al., 2004; Whiting, 1999).

Even though both implementation factors and system success have received a fair amount of attention in the data warehousing literature, as mentioned previously, they rarely are examined in the same study. In other words, the connection between the two groups of variables (the arrow in Figure 1) remains unresolved (Mukherjee & D’Souza, 2003). Consequently, we seek to answer the research question: Which implementation factors have an effect on which system success variables? Given the state of the research, no formal hypotheses are developed; instead, a series of best-subset regressions are used to explore the relationship.

Conceptually, there could be interaction among implementation factors and among success variables (e.g., DeLone & McLean, 1992, 2003). Those relationships are beyond the scope of this research. Similarly, another group of variables, implementation success, could be added to the model to mediate the effect of implementation factors on system success (Wixom & Watson, 2001). Extensions to current research are discussed later.

Methodology

Data Collection

Based on the research model discussed above, a web-based questionnaire was developed to collect data on the eight data warehousing success variables and eleven implementation factors as 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, & Watson, 1988). This approach is not without its limitations; however, such as the possible interaction of the factors (Nandhakumar, 1996) and the exclusion of context variables including social, cultural, political, and economic factors. (Bussen & 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 largest group of respondents (38%) was data warehousing specialists, followed by data warehousing database administrators (21%). The remaining respondents were data warehousing managers (18%), data warehousing consultants (16%), and data warehousing 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.

The respondents were asked to rate how significant each of the success variables was using a five-point scale. They were also asked to rate how important each of the implementation factors was using a five-point scale.

Data Analysis

Best subsets regression allows for the identification of a limited number of variables that can produce the best prediction result. Many criteria have been proposed as the basis for the selection of the best subset of variables (see, for example, Hocking, 1976). Common criteria used in popular statistical packages such as Minitab include R^2 , adjusted R^2 , C_p , the total squared error, and S , the average prediction mean squared error. Generally, the larger the R^2 and adjusted R^2 , and the smaller the C_p and S , the better. Moreover, C_p should be close to P , the number of parameters (including the intercept) to be estimated (Draper and Smith, 1998), which ensures that the bias in the estimated parameters is small (Hocking, 1976). In addition, C_p should be no larger than P if the objective is to use the model for prediction (Hocking, 1976). Basically, as P increases, R^2 increases and C_p , which represents error, decreases. However, unlike R^2 , which always increases as more variables are included into a model, C_p will decrease initially, but start to increase when less relevant variables are included. The last rule aims to include as many variables as possible, so long as C_p is still in the decreasing stages. The resulting model is free of multicollinearity because variables are added only when they significantly contribute to the variance explained by the model, as manifested by a substantial increase in R^2 or decrease in C_p . As an added assurance, the variance inflation factor (VIF) may be calculated to detect the existence of multicollinearity. A general rule is that VIF should not exceed 10 (Belsley, Kuh, & Welsch, 1980).

Findings

Table 2 shows the eleven best subsets regression results on the first success variable, “easy to use.” For each possible number of variables to be used as predictors, Minitab calculates the four criteria mentioned above and displays the best model. The model with six variables was considered the best because its C_p is close to and less than P . A separate regression was then run using those six selected variables as predictors. The coefficients for the intercept and the factors, along with the R^2 value are displayed in Table 3. Similar procedures were followed to select variables and obtain regression results for other success variables. VIF for all the models was calculated. The largest VIF was 1.49, well within the acceptable threshold of 10.

Table 2: Best Subsets Regression on Easy to Use

	1 Factor model	2 Factor model	3 Factor model	4 Factor model	5 Factor model	6 Factor model	7 Factor model	8 Factor model	9 Factor model	10 Factor model	11 Factor model
F1: clearly defined business needs/benefits					X	X	X	X	X	X	X
F2: top management support											X
F3: user involvement/participation	X	X	X	X	X	X	X	X	X	X	X
F4: source data quality				X	X	X	X	X	X	X	X
F5: proper development technology								X	X	X	X
F6: adequate IS staff and consultants									X	X	X
F7: project management (teamwork)			X	X	X	X	X	X	X	X	X
F8: practical implementation schedule						X	X	X	X	X	X
F9: proper planning / scoping of project							X	X	X	X	X
F10: adequate funding				X	X	X	X	X	X	X	X
F11: measurable business benefits		X	X							X	X
R ²	16.3	22.7	26.1	30.9	33.8	36.5	37.3	37.8	38.4	38.7	38.8
Adjusted R ²	15.4	21.1	23.8	28.0	30.2	32.3	32.4	32.2	32.1	31.6	31.0
C-p	23.6	16.6	13.8	9.1	7.1	5.2	6.2	7.4	8.6	10.2	12.0
S	.855	.826	.812	.789	.777	.765	.765	.766	.767	.769	.773

As shown in Table 3, all but one factor, “top management support,” is included in some best subsets models. At the same time, each success variable is a function of different implementation factors. All the factors in the best subsets models are considered critical except the ones with a negative coefficient. A factor with a negative coefficient means lower values (low importance) of this factor are associated with higher values (high significance) of a particular success variable; therefore, the factor may be statistically significant; in practice, it is anything but critical.

All the critical factors are shaded in Table 3 for easy identification. In addition, the regression coefficients represent the magnitude of the impact of each factor. While some coefficients are very close, others differ greatly. For example, for “better decisions,” the coefficient for “project management/teamwork” is more than three times larger than that for the other factors and, therefore, deserves special attention. On the other hand, given the exploratory nature of this study, findings should be interpreted as preliminary and subject to further validation and research.

Table 3: Regression Models of Data Warehousing Implementation Factors

Data Warehousing Success (Dependent Variable)	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	Intercept	R2
Easy to use	.192 (.035)		.225 (.027)	.288 (.003)			-.443 (.000)	.210 (.051)		.210 (.042)		1.022 (.057)	.365 (.000)
Speedy information retrieval			.243 (.014)		.294 (.007)				-.208 (.066)	.185 (.062)		1.997 (.000)	.212 (.000)
More information							.241 (.007)			.299 (.000)		1.491 (.000)	.232 (.000)
Better quality information				.376 (.000)			.276 (.005)					1.449 (.000)	.346 (.000)
Improved productivity			.215 (.010)			.147 (.069)				.213 (.019)	.203 (.010)	1.049 (.014)	.369 (.000)
Better decisions	.138 (.098)					.151 (.090)	.510 (.000)	-.155 (.101)			.201 (.014)	.696 (.142)	.447 (.000)
Improved business processes										.199 (.045)	.391 (.000)	1.561 (.000)	.279 (.000)
Increased competitive position	.205 (.050)						.342 (.006)		-.196 (.107)		.558 (.000)	.097 (.862)	.418 (.000)
¹ Independent variables for the regression models are:													
F1: clearly defined business needs/benefits						F7: project management (teamwork)							
F2: top management support						F8: practical implementation schedule							
F3: user involvement/participation						F9: proper planning / scoping of project							
F4: source data quality						F10: adequate funding							
F5: proper development technology						F11: measurable business benefits							
F6: adequate IS staff and consultants													
² The first number in each cell (except the last column) is the regression coefficient; the second number in parenthesis is the p-value													

Five factors are considered important to the perception of how easy a data warehouse is to use: “clearly defined business needs/benefits,” “user involvement/participation,” “source data quality,” “practical implementation schedule,” and “adequate funding.” For a data warehouse that offers “speedy information retrieval,” “user involvement/participation,” “proper development technology,” and “adequate funding” are critical. Wixom and Watson (2001) have shown that project implementation success positively affects system quality. To the extent that the above factors contribute to implementation success, it is not surprising that they have a positive effect on the two system quality variables: “easy to use” and “speedy information retrieval.”

“More information” is a function of two critical success factors: “project management” and “adequate funding.” “Project management,” along with “source data quality,” is critical to the production of “better quality information.” Both “source data quality” and “project management” are related to the technical aspect of a project. Wixom and Watson (2001) found that technical

factors have a positive impact on technical implementation success, which in turn has a positive (although nonsignificant) effect on information quality.

A variety of factors are responsible for a data warehouse's individual benefits. To achieve "improved productivity," "user involvement/participation," "adequate IS staff and consultants," "adequate funding," and "measurable business benefits" are critical. The benefit of "better decisions," however, is dependent on "clearly defined business needs/benefits," "adequate IS staff and consultants," "project management," and "measurable business benefits."

Only two factors are critical to "improved business processes": "adequate funding" and "measurable business benefits." Finally, a data warehouse may enhance a company's competitive position if three factors are in place: "clearly defined business needs/benefits," "project management," and "measurable business benefits."

Since not much research has been done to investigate the determinants of a data warehouse's individual or organizational benefits, interpretation of our initial evidence warrants caution. Nevertheless, the factor that is the most common among all regression models and thus maybe the most important is "measurable business benefits." This makes sense since in order to produce tangible benefits such as "improved productivity" or "better business process," the deliverables must be defined beforehand. Even though ROI analysis has not been regularly done in the past, more companies are taking the steps to define and measure business benefits from their warehouse (Sinn, 2003; Watson et al. 2004; Whitting, 1999). Support for this factor is succinctly summarized in the quote: "you can't manage what you can't measure." (Sinn, 2003, p. 90).

Discussion and Implications

Limitations

Several limitations to the current study have been mentioned; notably, the research model could be expanded or modified to investigate additional relationships. One example would be the testing of relationships among success variables, following the line of research in IS success model (DeLone & McLean, 1992, 2003).

Another limitation is that all the survey respondents are data warehousing professionals. While they are in a good position to evaluate the importance of various implementation factors, their perception on the success of the delivered system is probably not without bias. Future research should include the views of end users of data warehouses. More research opportunities are highlighted in the remainder of the paper.

Practical Implications

This research has contributed to the understanding of factors that impact data warehousing success. By examining success variables individually, the factors that are important and the strength of their effects on different success variables are identified. The magnitude of the effect of each implementation factor, as represented by the standardized regression coefficient in Table 3, provides a handy reference for practitioners' planning and implementation of data warehouses. Unlike existing guidelines found in the literature, which usually list several factors without being specific about their effects on data warehousing success, these figures reveal the critical success factors and their effects in unambiguous terms. Since different organizations may have different objectives or different "data warehousing sweet spots" (Watson et al., 2004, p. 7), these figures are useful in identifying the exact factors that need attention and in providing a basis for prioritizing those factors if need be, as explained next.

For companies that are relatively new to data warehousing, the focus may be on the ease of use of the system. If so, according to the first regression model in Table 3, it is advisable to concentrate on five success factors: “source data quality,” “user involvement/participation,” “adequate funding,” “clearly defined business needs/benefits” and “practical implementation schedule,” in that order of importance. Since greater user involvement/participation facilitates the incorporation of user input into systems design, the resulting system is likely to be easy to use. The linkage between other factors and ease of use is not as apparent and may need further study; however, it is prudent to be heedful of their effects. On the other hand, for organizations that are more interested in using data warehousing for competitive advantage, “measurable business benefits” is paramount, followed by “project management/teamwork,” and “clearly defined business needs/benefits” (see the last regression model in Table 3). It is obvious that “user involvement/participation” may be important to building a warehouse that is “easy to use.” It is more important to make a business case for the warehouse (e.g., “measurable business benefits” and “clearly defined business needs/benefits”) if the intent is for strategic applications. In short, our results provide an empirically validated list of critical success factors that can be tailored to suit the needs of individual companies.

Research Implications

Columns in Table 3 provide an indication of how “critical” each implementation factor is. Overall, “top management support” is the only factor found insignificant. 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 this factor 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 & 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. More research into this factor’s impact on different success variables seems warranted.

Factors four to seven are related to technical aspects of data warehousing. “Project management/teamwork” is especially noteworthy as it affects all but the most basic success variable, system quality. It seems that, as is true in all large-scaled system development projects, the most critical “technical” factor is not the people, nor the technologies, but the way that all the resources are pooled, organized, and managed. This factor may not have a direct effect on how the completed system is perceived, but it most likely will affect how a system is used and how individual and organizational benefits are derived from its usage. The significance of project management is echoed by many practitioners as exemplified by the comment that project management skills are “the single most important technology resource that comes to play on data warehouse development” (Trembly, 2001, p. 41).

Factors eight and nine are related to the scheduling aspect of a project, and they are found to be less important than other factors. The only significant factor is a “practical implementation schedule,” which is needed for a warehouse to be “easy to use.” More research may be needed to better identify and measure schedule factors.

The final two factors are related to the economic aspects and both factors are very critical to the success of a data warehouse implementation. Data warehousing is very expensive and may be getting more so as companies are spending more in a race to build ever-larger warehouses. In addition to proper funding, “measurable business benefits” are critical for reaping tangible

benefits that will manifest in better performance at both the individual employees' level and the firm's level. This finding is consistent with the recent calls to pay more attention to data warehouse ROI (Lewis, 2001; Sinn, 2003; Watson et al., 2004; Whiting, 1999).

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 factors or success variables, 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 an additional success variable related to system quality.

As mentioned earlier, most prior studies have examined either implementation 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. They could choose to focus on one particular factor, e.g., a specially designed query processor, and study its effect on different success variables. They could also extend the current research by examining which factors are significant for a new benefit, e.g., "easy to manage." We believe that the list of critical factors for different benefits 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 & Myers, 1997) should be further researched to allow a fuller understanding of data warehousing success.

References

- Agosta, L. (2004, November). Data warehousing lessons learned: A time of growth for data warehousing. *DM Review Magazine*. Retrieved on 1/17/2005, from http://www.dmreview.com/article_sub.cfm?articleId=1012461
- Baker, S. & Baker, K. (1999). The best little warehouse in business. *Journal of Business Strategy*, 20(2), 32-37.
- Belsley, D., Kuh, E., & Welsch, R. (1980). *Regression diagnosis: Identifying influential data and sources of collinearity*. New York: John Wiley & Sons.
- Bussen, W., & 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., & Frolick, M. (2000). Measuring user satisfaction with data warehouses: An exploratory study. *Information & Management*, 37, 103-110.
- Conner, D. (2003). Data warehouse failures commonplace. *Network World*, 1(3), 24.

- Cooper, B.L., Watson, H.J., Wixom, B.H., & Goodhue, D.L. (2000). Data warehousing supports corporate strategy at First American Corporation. *MIS Quarterly*, 24(4), 547-567.
- DeLone, W. & McLean, E. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research*, 3(1), 60-95.
- DeLone, W. & 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. & 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
- Draper, N.R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). New York: John Wiley & Sons.
- Gagnon, G. (1999, March 19). Data warehousing: An overview. *PC Magazine*, 245-246.
- Gardner, S. (1998). Building the data warehouse. *Communications of the ACM*, 41(9), 52-60.
- Gorla, N. (2003). Features to consider in a data warehousing system. *Communications of the ACM*, 46(11), 111-115.
- Heun, C.T. (2000, December 11). Harrah's bet on IT to understand its customers. *InformationWeek*, 816, 10-12.
- Hocking, R. (1976). The analysis and selection of variables in linear regression. *Biometrics*, 32(1), 1-49.
- Hwang, M.I., & Cappel, J.J. (2002). Data warehouse development and management: Practices of some large companies. *Journal of Computer Information Systems*, 43(1), 3-6.
- Hwang, M.I., & Thorn, R. (1999). The effect of user engagement on system success: a meta-analytical integration of research findings. *Information & Management*, 35(4), 229-236.
- Johnson, L.K. (2004). Strategies for data warehousing. *MIT Sloan Management Review*, 45(3), 9.
- Joshi, K. & Curtis, M. (1999). Issues in building a successful data warehouse. *Information Strategy*, 15(2), 28-35.
- Kelly, S. (1997). *Data warehousing in action*. Chichester: John Wiley & Sons.
- 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. (2004, December 13). Too much information. *Forbes*, 110-115.
- Magal, S. R., Carr, H. H., & Watson, H. J. (1988). Critical success factors for information center managers. *MIS Quarterly*, 12(3), 413-425.
- Mukherjee, D., & 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(1), 62-72.
- Sammon, D., & Finnegan, P. (2000 Fall). The ten commandments of data warehousing. *Database for Advances in Information Systems*, 31(4), 82-91.
- Sakaguchi, T. & Frolick, M. (1997). A review of the data warehousing literature. *Journal of Data Warehousing*, 2(1), 34-54.
- Shim, J.P., Warkentin, M., Courtney, J.F., Power, D.J., Sharda, R., & 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.

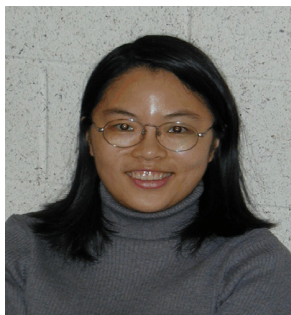
Data Warehousing Success

- 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.
- 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., & 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.
- Watson, H., Abraham, D., Chen, D., Preston, D., & Thomas, D. (2004). Data warehousing ROI: Justifying and assessing a data warehouse. *Business Intelligence Journal*, Spring, 6-17.
- Watson, H., Annino, D., Wixom, B., Avery, K. & 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., & Haley, B. (1997). Data warehousing: A framework and survey of current practices. *Journal of Data Warehousing*, 2(1), 10-17.
- Watson, H., & Haley, B. (1998). Managerial considerations. *Communications of the ACM*, 41(9), 32-37.
- Whiting, R. (1999, May 24). Warehouse ROI. *InformationWeek*, 735, 99-104.
- Winter, R., & 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. & Watson, H.J. (2001). An empirical investigation of the factors affecting data warehousing success. *MIS Quarterly*, 25(1), 17-41.

Biographies



Mark I. Hwang is a Professor of Management Information Systems. He holds a Ph.D. in Business Computer Information Systems from the University of North Texas. Dr. Hwang has published articles in business and information systems journals including *Advances in Accounting*, *Business Intelligence Journal*, *Data Base, Information & Management*, *Information Resources Management Journal*, *International Journal of Business*, *Journal of Computer Information Systems*, *Journal of Information Science*, *Journal of Information Technology Management*, *Journal of Management Systems*, *Managerial Auditing Journal*, *Multinational Business Review*, and *Omega*. His research interests include data warehousing, data mining, and meta-analysis.



Hongjiang Xu is an Assistant Professor of Business Information Systems Department at Central Michigan University. She received her PhD from the University of Southern Queensland, Australia. Her research interests include data and information quality, e-Commerce, data warehouse, and database management. Her work has been published in journals such as *Journal of Database Management*, *the Journal of Computer Information Systems*, *Advances in Management Information Systems*, *Industrial Management & Data Systems*.