ABSTRACT

Corporations are increasingly turning to data warehouses for support of critical corporate decision-making; however, much remains to be learned about the success of data warehousing. Building a data warehouse is no small endeavor, requiring considerable resources from the corporation in terms of both time and money. Ensuring the success of the system will generate for the firm a return on its considerable investment.

The research presented in this article investigated success in data warehousing as measured by user satisfaction. Data warehouse users at an online travel company were asked to complete a quantitative survey exploring their satisfaction with the company’s data warehouse system. Individual differences were explored to determine what variables influenced the satisfaction of users of differing education levels, computer expertise, data warehousing expertise, job classification, and tenure with the firm. Results of the research will assist organizations in maximizing the potential returns from the significant investment required to build a data warehouse.

Keywords: Data Warehouse; User Satisfaction; System Success; Information Technology

INTRODUCTION

Corporations are increasingly dependent on information technology as a means of achieving and retaining competitive advantage. In many cases, information technology is no longer seen as a cost of doing business, but rather is being exploited to ensure the success and survival of the firm. In the increasingly complex business world, massive amounts of relevant internal and external data exist that, if captured and properly analyzed, will provide valuable guidance in the increasingly complex realm of management decision-making. And yet, modern organizations are still said to be drowning in data but starving for information.

Organizations have vast stores of historical operational data of which they are anxious to take advantage, including valuable data about many aspects of their current business, such as customers, products, inventories, and sales. Until recently, companies did not have enough computing power at their disposal to be able to take advantage of these stored data [20]. Now, however, parallel processing technology, which increases both the speed of data analysis and the volume of data that can be analyzed by dividing a complex task among multiple linked processors, has advanced and is enabling organizations to begin to analyze these massive amounts of historical data. Data warehouse systems have capitalized on this advance in technology.

Murtaza [25] and Moye and Upton [24] identified two reasons why corporations are turning to data warehouses: to optimize control of current operations or to gain significant competitive advantage. Johnson [19] discusses the value of data warehousing applications in linking supply-chain partners. “This melding of data promises enticing advantages, such as real-time updating of customer information, the opportunity to focus on core capabilities and instant forecasting of inventory needs” (p. 9). There are numerous strategic advantages available from data warehouse applications. Murtaza [25] describes this strategic nature by stating that a data warehouse is, “an integrated, nonvolatile catalog of organizational data that is convertible into actionable information for strategic decision making” (p. 21). Data warehousing has been compared to mining for gold and many experts argue that data warehousing is a necessity for organizations to survive within the hypercompetitive global marketplace [24, 29].

The Problem Background

While corporations are increasingly turning to data warehouses for support of critical corporate decision-making, there remains much that is unknown about the success of data warehousing [29]. One study of small to mid-sized US firms found that firms desiring to implement a data warehousing solution, or any major infrastructure type innovation, should pay careful attention to both organizational and technology factors in order to promote success [29]. Building a data warehouse is no small endeavor, requiring considerable resources from the corporation in terms of both time and money. Estimates on the cost of establishing a data warehouse vary among authors, but range from $50,000 to several million [13, 22, 29, 36]. Potential returns are also high, with return on investment
Various researchers have measured the success of a data warehouse system is needed. Unfortunately, data warehouse projects also abound. For example, a major regional bank located in the Southeast United States lost $60 million in 1990. A new management team was installed and initiated a customer intimacy strategy based on a data warehouse [35]. Using the data contained in the system, the firm was able to determine the profitability of each client and product; to develop programs to attract, maintain, and enhance their customer base; to create profitable new products and services; and to redesign their distribution channels to maximize efficiency, profitability and customer satisfaction. The application of data warehouse technology helped the firm to return to a profitable position and achieve the status of innovative leader in the financial services industry. Similarly, data warehousing was found to be a valuable tool for the banking industry in Taiwan [15] and Korean financial companies [14].

Another example of a successful data warehousing application comes from the major manufacturer Whirlpool, which uses the technology to assure quality in its home appliances [35]. Whirlpool’s data warehouse is used to track and analyze everything associated with the appliances that they manufacture, starting with the components purchased from suppliers and continuing through customers’ life-long experiences with the products. This accumulation of detailed data allows quality engineers to track the performance of component parts, purchasing agents to find the lowest-cost, highest-quality part available on a global scale, and even suppliers to assess the performance of parts supplied to Whirlpool.

Unfortunately, data warehouse projects also have high failure rates, averaging 50% [5, 29]. Many firms fail to realize desired benefits of better customer service or reduced production and inventory costs [19], with 85% of all data warehouse projects failing to meet their intended objectives, and 40% not advancing beyond the planning phase [12]. This might be attributed to the significant staff, hardware, software, time and funds needed to implement such projects [15, 2]. Other reasons include the poor quality of data in the warehouse [19, 10], lack of executive sponsorship [15, 6, 33], too many different technologies [32], unrealistic expectations [28, 34], and weak scalability and compromised decision support [30].

Given that the literature indicates that potential returns from a data warehouse are high, but that the threat of failure is also high, more understanding of the factors that contribute to the success of a data warehouse system is needed. Various researchers have measured the success of information systems in organizations, identifying six major categories of information system (IS) success: system quality, information quality, use, user satisfaction, individual impact, and organizational impact [7]. However, user satisfaction is considered by many to be the most useful measurement of system success [4].

Since the inception of the technology, data warehousing has been considered a valuable tool for managerial decision-making, but little academic research exists to demonstrate the success of the technology in organizations. Despite the importance placed on data warehouse systems, only one empirical study has been conducted to measure system success of these systems as defined by user satisfaction [4]. There is a need for academic investigation into the success of data warehouse projects. This research will address that need through an empirical investigation of success in data warehousing as measured by user satisfaction.

**LITERATURE REVIEW**

The literature review presented here focuses on two primary areas: understanding data warehouse technology and measurement of user satisfaction in computer-based information systems.

**Data Warehouse Defined**

The recent dramatic increases in processing power and disk storage capacities have allowed corporations to consider information technology opportunities that were previously unavailable, due to the limited state of technology, or unfeasible due to cost considerations [38]. The emergence of parallel processing capabilities underlies the development of data warehousing, which has been called the center of the architecture for information systems of the future [16]. McFadden, Hoffer and Prescott define a data warehouse as “An integrated decision support database whose content is derived from the various operational databases” [23, p. 602]. Levine and Siegel [20] define data warehousing as a group of technologies that integrates the operational data of all departments into a single database, which is arranged by subject, designed to support high level analysis and decision making. Inmon [16], who is widely regarded as the father of data warehousing, defines a data warehouse as a subject-oriented, integrated, time-variant, nonvolatile collection of data used in support of management decision-making processes. These foundation concepts are discussed in greater detail in the following paragraphs.

Subject orientation refers to the basic data structure of a data warehouse [20]. A data warehouse is organized around the key subjects (or high-level
entities) of the enterprise, rather than around the business function. Examples of major subjects include customers, patients, students, and products.

The data in a data warehouse are integrated from disparate source systems. In order to achieve full integration, the data must be fundamentally altered to share commonality in definition, layout, relationships, etc. [20]. The data housed in the data warehouse are defined using consistent naming conventions, formats, encoding structures, and related characteristics [17]. Firms wishing to integrate their data with one or more outside partners must carefully consider all internal business processes to ensure that existing data are in a format on which managers and employees can rely [19].

The fundamental operation of the data warehouse is to process historical data into valuable current and future information. Therefore, data in the warehouse contain a time dimension in order that they may be used as a historical record of the business [17, 20].

Nonvolatility of data refers to the stable nature of the data warehouse. This stability is achieved by prohibiting real-time updating of records. Rather, data in the data warehouse are loaded and then refreshed as needed from operational systems, but cannot be updated by end users [27, 17]. Although nonvolatility and limited update have been fundamental concepts since the inception of the technology, experts suggest that there is a strong move toward the incorporation of more real-time data [19, 26] through more frequent updating of the data warehouse from operational sources.

User Satisfaction as a Measure of System Success

Multiple researchers have been involved in the quest for a valid measure of user satisfaction. The original efforts in this area focused on user satisfaction in traditional information systems. That is, systems in which users do not directly interact with the system, but rather interface with systems professionals and receive only the output of the systems. Bailey and Pearson [1] conducted the earliest research. They developed a semantic differential instrument containing 39 items measuring overall computer user satisfaction. These researchers conducted interviews with middle managers to determine what factors were most important to user satisfaction with information systems. Ives, Olsen, and Baroudi [18] conducted a factor analysis of the Bailey and Pearson [1] instrument and, based on the results of their research, proposed a 13-item scale.

Doll and Torkzadeh [8] later investigated the measurement of user satisfaction in end-user computer environments. Unlike the more traditional computing environment, this environment is one in which system users interface directly with the information systems. Drawing on the work of Bailey and Pearson [1] and Ives, Olsen, and Baroudi [18], these researchers generated 31 items to measure end-user perceptions. Additional questions covering the construct of ease-of-use of the system were added. These researchers developed a 40-item instrument using a five-point Likert scale. Based on the results of the statistical analysis, the researchers published a revised instrument [8] and it was then validated as a 12-item scale [9].

The most recent work in this area was conducted by Shaw, Lee-Partidge and Ang [31]. The researchers examined satisfied and dissatisfied end-users in an organization to determine if the users hold different technological frames of reference toward end-user computing. The results of the study show that frames of reference affected the users' satisfaction and therefore their expectations of the technology, their interactions with support personnel, and their willingness to rely on the technology on a daily basis.

Mahmood, Burn, Leopoldo and Jacquez [21] conducted a meta-analysis of research results in 45 end-user satisfaction studies. The researchers found wide variability in factors reported to affect end-user satisfaction. The factors reported most frequently in the empirical literature included: End-user participation in the design of the system, technical support provided to the end-user, management support, previous computing experience for the end-user, and end-user computing attitudes.

The only study to investigate user satisfaction with data warehouse systems focused on validating an instrument for this genre of information systems. The researchers utilized a 21-item scale that was developed from the Doll and Torkzadeh [8] instrument [4]. The researchers added items that investigated the importance of the support provided to the data warehouse user by the corporate information center. An exploratory factor analysis resulted in 16 retained items loading on three factors: Support provided to the end user; Accuracy, format and preciseness of data; and Need fulfillment. The authors concluded that support provided by the corporate information center was important to the satisfaction of the users of the data warehouse. The study was limited by the small sample of 42 participants, but resulted in a tentative declaration that the tools developed to measure user satisfaction in other information systems seemed to be valid instruments for similar investigations of data warehouse systems.

Hong, Katerattanakul, Hong, and Cao [14] conducted a study related to user satisfaction. The researchers investigated factors contributing to end-user usage of a data warehouse. Data for the study
was gathered in Korean financial companies. The results suggested that data quality and end-user support and training were the most significant system characteristics affecting end-users’ usage of the data warehouse.

Ford, Ledbetter, and Roberts [11] discovered that demographic factors, prior computer training and experience had significant impact on the participants’ use of computers. Given that these individual differences were seen to affect computer usage, intuition would suggest that satisfaction with a system, in this case a data warehouse, may also be affected.

**Purpose of the Study**

The purpose of this study was to investigate user satisfaction in data warehousing. Specifically, the researcher desired to identify variables contributing to system success as measured by user satisfaction. Individual differences were explored to determine what variables influence the satisfaction of users of differing education levels, computer expertise, data warehousing expertise, job classification, and tenure with the firm.

**METHODOLOGY**

A major firm in the on-line travel industry agreed to participate in the study. The company pioneered the online travel services, and continues to be the most popular travel service on the Web, giving consumers access to hundreds of airlines, thousands of hotels and cruise, last-minute and vacations packages, and best-in-class car rental companies, all backed by 1,000 customer service representatives staffed to provide 24-hour assistance. The subjects, all data warehouse users for the participating firm, were informed that the study was part of a scholarly project and that participation was voluntary. As a result, any person who did not wish to participate was excluded from the study.

The company listed 162 registered users of the data warehouse. All registered users were sent an email inviting their participation in the study. In the email, the purpose of the study was clearly stated and confidentiality of responses was guaranteed. The email contained a link to the URL of the survey. The survey was hosted on a website run by an online survey company retained by the researcher. After three days, non-respondents were sent another request for survey participation. After four more days, for a total of seven days, the data gathering was closed. Ninety-one responses were obtained, yielding a response rate of 56%. Upon examination, six surveys were determined to be incomplete and therefore unusable. One additional survey contained multiple responses for some questions and was not used. The final number of usable surveys was 84, for a response rate of 52%.

The survey instrument developed by Chen, Soliman, Mao, and Frolick [4] served as the basis for the research. The instrument was modified to seek information on the sophistication of the user by determining their years of general computer experience and years of data warehouse system experience and to gather demographic data about the user such as gender, level of education, and position within the firm.

**RESULTS**

The majority of the respondents indicated their position to be either functional department staff or functional department management (about 67%) and employed with the company for 2-5 years (55%). The level of education among the respondents was high, with 37% holding bachelor’s degrees and 35% holding graduate degrees. The expertise of the respondents when using the data warehouse system was evenly distributed, while general computer expertise was rated experienced (77%).
The majority of respondents (73%) indicated that they were somewhat successful in obtaining needed information from the data warehouse. When asked how vital the information obtained from the data warehouse had been either in performing their job better or in transforming their job, 67% of respondents rated the importance as somewhat important, while 33% rated the importance as vital.

Data Analysis

For data analysis, it was determined that the survey responses provided by those working in the information systems function within the company should be excluded. As stated, the purpose of the research was to determine the level of end-user satisfaction with the data warehouse application. Because the IS personnel are not classified as end-users, their responses were removed from the data analysis. Therefore, an additional thirteen (13) surveys were eliminated, leaving 71 end-user surveys for analysis.

Hypothesis 1

The null hypothesis for hypothesis one stated that there is no significant difference in user satisfaction among users with various levels of education. The results of the ANOVA test show that, at the .05 level of confidence, there is a significant difference in user satisfaction among users with various levels of education.

However, the researcher expected that users with higher levels of education would experience more satisfaction with the data warehouse. The results of the data analysis indicate that users with higher levels of education were actually less satisfied with the data warehouse.

The ANOVA test of significance determined that, at a .05 level of confidence, there is a statistically significant difference in the mean level of satisfaction as reported by users with differing levels of education. The calculated test statistic is .039 (F=2.498, p<.05), supporting the determination of a significant difference. Therefore, the null hypothesis is rejected.

Hypothesis 2

The null hypothesis for hypothesis two stated that there is no significant difference in user satisfaction among users from various positions within the firm. Alternatively, the researcher hypothesized that users at higher levels within the firm will experience more satisfaction with the data warehouse.

The results of the ANOVA test for significance, at the .05 level of confidence, determined that the level of significance in the reported levels of satisfaction among users with differing levels of tenure with the company was .564 (F=.686). Therefore, the null hypothesis was accepted. No significant difference was noted in user satisfaction among users with differing lengths of tenure with the company.

Hypothesis 3

The null hypothesis for hypothesis three stated that there is no significant difference in user satisfaction among users with differing lengths of tenure with the company. Alternatively, the researcher hypothesized that users with longer lengths of tenure with the company will experience more satisfaction with the data warehouse.

The results of the ANOVA test for significance, at the .05 level of confidence, determined that the level of significance in the reported levels of satisfaction among users with differing lengths of tenure with the company was .564 (F=.686). Therefore, the null hypothesis was accepted. No significant difference was noted in user satisfaction among users with differing lengths of tenure with the company.

Hypothesis 4

The null hypothesis for hypothesis four stated that there is no significant difference in user satisfaction among users with differing levels of general computer expertise. Alternatively, the researcher hypothesized that users with higher levels of general computer expertise would experience more satisfaction with the data warehouse.

This hypothesis was tested through two questions on the survey. First, users were asked to indicate the number of years’ experience they had with computer software. The results of the ANOVA for years of experience with computer software, at the .05 level of confidence, determined that the level of significance in the difference of reported levels of satisfaction among users with differing years’ experience with computer software was .559 (F=.586). Therefore, the null hypothesis was accepted. No significant difference was noted in user satisfaction among users with differing lengths of tenure with the company.

Hypothesis four was also tested through another question on the survey. The question asked users to self-report general computer expertise. The results of the ANOVA test for significance, at the .05 level of confidence, determined that the level of significance in the reported levels of satisfaction among users with differing levels of self-reported general computer expertise was .154 (F=1.925). Based on the results of both questions, the null hypothesis for hypothesis four was accepted. No significant differences were noted in user satisfaction among users with differing levels of general computer expertise.

Hypothesis 5

The null hypothesis for hypothesis five stated that there is no significant difference in user satisfaction among users with differing levels of data warehousing expertise. Alternatively, the researcher
expected that users with higher levels of data warehousing expertise would experience more satisfaction with the data warehouse.

This hypothesis was tested through two questions on the survey. First, users were asked to indicate the number of years’ experience they had with the data warehouse system. The results of the ANOVA for years of experience with the data warehouse, at the .05 level of confidence, determined that the level of significance in the difference of reported levels of satisfaction among users with differing years’ experience with data warehouse was .955 (F=.107).

Hypothesis five was also tested with another question on the quantitative survey. The question asked users to self-report their data warehouse expertise. The results of the ANOVA test for significance, at the .05 level of confidence, determined that the level of significance in the reported levels of satisfaction among users with differing levels of self-reported data warehouse expertise was .043 (F=2.858, p=.05), a statistically significant result. Based on the results of this question, the null hypothesis for hypothesis five was rejected. There was a significant difference in user satisfaction among users with differing levels of data warehouse expertise.

However, the researcher expected that those users who self-reported high levels of data warehouse expertise would experience higher levels of satisfaction with the system. Based on the analysis of the data, that was not the case. Satisfaction appears to be highest for users at the beginner level, slumps in the novice category, rebounds with the experienced users, and is lowest with the expert users.

**Research Hypothesis 6**

The null hypothesis for hypothesis six stated that there is no significant difference in user satisfaction among users with differing levels of individual success in accessing information contained within the data warehouse. Alternatively, the researcher hypothesized that users with higher levels of individual success in accessing information contained within the data warehouse will experience more satisfaction with the data warehouse.

The results of the ANOVA test for significance, at the .05 level of confidence, determined that the level of significance in the reported levels of satisfaction among users with differing levels of individual success in accessing information contained within the data warehouse was .089 (F=2.511). Therefore, the null hypothesis was accepted. No significant difference was noted in user satisfaction among users with differing levels of individual success in accessing information contained within the data warehouse.

**Discussion**

Research question one sought to determine if individual differences in educational level affect user satisfaction with the data warehouse system. The researcher suspected that higher levels of education would lead to higher levels of satisfaction with the data warehouse application. However, the results of the study suggested that the higher a users’ education level, the lower the level of satisfaction with the data warehouse. This may be due to users with higher levels of education being more knowledgeable about system requirements and more demanding in relationship to data quality. Users with higher levels of education may be less likely to take the results of a query at face value and may be more likely to ask questions about system functionality.

Research question two sought to determine if individual differences in position within the company affected user satisfaction with the data warehouse system. The researcher suspected that users serving at higher organizational levels within the firm would experience more satisfaction with the data warehouse. This research question was based on the assumption that data warehouse applications are strategic applications and that users at higher levels within the firm would have more knowledge concerning the strategy of the firm. The results here suggested that there is no difference in satisfaction among users at various levels within the firm. This result may suggest that users at all levels of the organization are able to grasp the importance of the system and utilize it in their job function.

Research question three sought to determine if individual differences in tenure with the company affected user satisfaction with the data warehouse system. The researcher suspected that users with longer lengths of tenure with the organization would experience more satisfaction with the data warehouse. This research question proposed that data warehouse applications are strategic applications and users with longer tenures with the firm would have more knowledge concerning the strategy of the firm. The results of the research found no difference in satisfaction among users with various lengths of tenure with the firm. This result suggests that all users, regardless of the length of their tenure, are nearly equally satisfied with the data warehouse.

Research question four sought to determine if individual differences in general computer expertise affected user satisfaction with the data warehouse system. It was suspected here that those users with higher levels of expertise using computer software would be more satisfied with the data warehouse. Ford, Ledbetter, and Roberts [11] discovered that demographic factors, prior computer training and experience had significant impacts on
the participants’ use of computers. Given that these individual differences were seen to affect computer usage, intuition would suggest that satisfaction with a system, in this case a data warehouse, may also be affected. However, the results of this study suggested that general computer expertise did not seem to affect user satisfaction with the data warehouse.

Research question five examined individual differences in expertise with the data warehousing application and user satisfaction with the system. It was expected that users with the highest levels of expertise would report the highest satisfaction levels. Users with varying levels of self-reported data warehouse expertise did report a statistically significant difference in the level of satisfaction. Satisfaction appeared to be highest for users at the beginner level, it tended to slump in the novice category, rebounded with the experienced users, and was at its lowest with the expert users.

These differences may be explained by two factors: 1) the learning curve inherent in mastering the data warehouse application, and 2) issues of data quality and support. Beginning users may be the most satisfied with data warehouse tool due to the initial impression made by a very powerful system. Those in the novice category may be less satisfied because they are struggling to master some of the intricacies of system and may become frustrated by the steep learning curve. Experienced users, having survived the learning curve, may become more satisfied than those in the novice category due to the satisfaction brought by success in mastering the system. Finally, expert users may be the least satisfied because they are also the most demanding of the system. They may continually want more and may be very critical of any data quality issues.

Research question six investigated the impact of individual success in accessing the data warehouse on user satisfaction. The results of the study suggested that there are no statistically significant differences in the level of satisfaction reported by system users who have varying levels of individual success with the data warehouse. This result may be best explained by looking at the survey responses. When responding to the individual success survey question (Overall, how successful would you say you have been in accessing needed information from the data warehouse?), only two (2) respondents indicated that they were not at all successful (mean satisfaction 2.5), while 56 percent indicated somewhat successful (mean satisfaction 3.02) and 13 indicated very successful (mean satisfaction 3.54). It would appear, based on these responses, that the users of the data warehouse system at the subject firm were at least somewhat successful in accessing the needed data. However, an examination of the mean scores on overall satisfaction of the users in these categories shows that there are higher levels of satisfaction reported by users with more individual success. A correlational analysis may be more appropriate here, with a larger sample.

**CONCLUSION**

Despite vast quantities of data collected in organizations today, many managers have difficulty obtaining the information they need for decision-making. This problem has given rise to the development of data warehouses. Data warehouses support managers’ complex decision making through applications such as the analysis of trends, target marketing, competitive analysis, and many others. Data warehousing has evolved to meet the needs of decision-makers without disturbing existing operational processing.

The process of developing a quality data warehouse is long and complicated; however, the potential returns to the firm are enormous. Despite this potential, only one previous study had been conducted to measure data warehousing system success as defined by user satisfaction. This research attempted to contribute to the academic literature by investigating the factors that contribute to this vital measure of system success. As discussed previously, building a data warehouse is a considerable investment for a firm. As such, firms want the data warehouse project to be a success and therefore must seek to maximize the satisfaction of their users who are accessing the data warehouse.

**Limitations**

The primary limitation in this research study arose from having only one company participating. This limited the ability to generalize the results beyond this setting. In addition, respondents were asked to self-report their computer expertise, as well as their data warehouse expertise. These self-reported values were not validated through any empirical testing and therefore biases could have resulted. The sample size was clearly limited also.

**Recommendations for Future Research**

For future research, this study could be repeated with a broader, larger sample. For example, the study could be repeated with users from multiple firms across multiple industries. Another possible scenario would be to repeat this study with multiple firms in the same industry and compare the results further. With either research design a larger sample would, hopefully, confirm the results of this study.
and would also uncover other statistically significant results.

Also, the impact of organization culture on system success as measured by user satisfaction must be probed. That variable was not specifically investigated here, and yet it was thought to be important variable as evidenced in the interview process.

There is a need for more research into the satisfaction of expert users. In this study, those users who classified themselves as experts with the data warehousing system were the least satisfied with the data warehouse. Although the researcher speculated on the cause of this result, more research is needed to confirm the results and to empirically determine the root of the dissatisfaction.

Finally, a suggestion for future research would be to conduct a longitudinal study on user satisfaction in data warehousing. The researcher should collect data on user satisfaction and then assist the firm in implementing some of the suggestions that derived from this research, i.e. a help desk, training classes, data warehouse newsletter, etc. Following the implementation of these changes, user satisfaction could be measured again and compared with the present results.

REFERENCES


