Journal of International Technology and Information Management

Volume 14 | Issue 4 Article 4

2005

Task-Technology Fit in Data Warehousing Environments: Analyzing the Factors that Affect Utilization

Richard V. McCarthy
Quinnipiac University

George F. Claffey
Central Connecticut State University

Task-Technology Fit in Data Warehousing Environments: Analyzing the Factors that Affect Utilization

Richard V. McCarthy
Quinnipiac University

George F. Claffey
Central Connecticut State University

ABSTRACT

Enterprise data warehouses are an expensive investment in both time and resources. The promise of data warehouses is improved decision-making through user empowerment and enablement of sophisticated decision support tools. Have organizations been able to achieve this goal? Task-technology fit defines a model that suggests that for a technology to be utilized, it must meet the needs of a user and provide features that support the fit of the requirements of the task; performance impacts will result. When an organization commits the time and resources necessary to develop an enterprise data warehouse, their expectation will be a high task-technology fit. This study extends prior task-technology fit research to provide an evaluation of task-technology fit in data warehouse applications. The focus of the study is to examine the factors that contribute to the task-technology fit specifically within data warehouse environments.

INTRODUCTION

New data warehouse tools that allow for the drill down into many layers of data have enabled complex analysis of information. In order to effectively make use of these tools, there must first be an infrastructure in place that supports an integrated set of data that can effectively serve as the basis for supplying a single source of information. Implementation of a data warehouse enables that single source. An enterprise data warehouse is a centralized store of summary and detail information from all relevant sources that is used to analyze a business by allowing for drill down analysis and ad hoc discovery from multiple user groups (Kimball, 1996). A data warehouse contains four characteristics (Corey et. al., 1999) (Subramanian, et. al., 1997):

1. subject oriented
2. non-volatile
3. time variant
4. integrated

Subject orientation enables users to determine not only how their business is performing, but also why. A data warehouse differs from an operational database in which most operational databases have a product orientation and are tuned to handle transactions that update the database. There is a temporal and granularity mismatch in comparing applications driven from a data warehouse to an on-line transaction processing system caused by the amount of detail that each application type focuses on (Jarke, et. al., 1999). The non-volatility of a data warehouse means that the data do not change between updates. This allows the data warehouse to be tuned for improving the performance of accessing information, since issues such as allowance for free space (for data growth) can be ignored. A time variant data warehouse presents data as of a single point in time. All relevant data stores that are utilized are synchronized as of a single point in time. An integrated data warehouse manages all of the data needed for a business consolidated in a single location. Relevant data sources may include external data as well as internal operational data. A data warehouse should be designed to challenge people's thinking, not reinforce it (Kimball, 1996). It should lead to asking further questions to analyze information.

The focus of this study is to determine empirically the factors that affect the usage of a data warehouse system. Once the factors have been identified, we then test to determine if there is a task-technology fit for data warehouse systems. A task-technology fit has been shown to lead to system utilization; therefore, we seek to provide an empirical measure of data warehouse system usage. Factors are categorized based upon whether they are task or individual characteristics. This study further extends prior task-technology fit evaluations by analyzing the correlation of each of the factors.

This paper presents an overview of the development of the data warehouse as a critical component of an information systems architecture. It provides the theoretical foundation of task-technology fit theory as well as an analysis of the prior task-technology fit studies. A detailed description of the research methodology is then presented with an analysis of the results of our findings.

DATA WAREHOUSE PROJECT GOALS

Most data warehouses have six fundamental goals (Segars and Grover, 1993):

1. The data warehouse provides access to organizational data, immediately, and with high performance.
2. The data are consistent. The organization should have a *single version of the truth*.
3. The data in the data warehouse can be separated and combined by every organizational measure.
4. The data warehouse provides a set of tools for query and analysis.
5. Only reliable and complete data is published within the data warehouse.
6. The quality of the data in the warehouse can be used to drive business process reengineering.

Data warehouse projects are either data centered or application centered (Watson and Haley, 1998). A data centric warehouse is based upon a data model that is independent of any application. It is designed to support a variety of user needs and a number of applications. The methodological approach to design a data centric warehouse involves data modeling with a group of business experts who are familiar with the different information views that are needed to support that business. This consists of a top down approach in producing specifications of information needs so as to not leave data behind (Subramanian, et. al., 1997). A mapping approach should be used to provide a structured approach to classification of data. Data centric warehouses should support flexibility because executive information needs change constantly based upon changes in the underlying business. An application centric warehouse is one that is initially designed to support a single initiative or small set of initiatives. It is expected that the task-technology fit will be higher for application centric data warehouses than for data centric data warehouses; this proposition must be validated.

In addition to flexibility, a data warehouse needs to support scalability. The main issues within scalability are the amount of data within the warehouse, the number of concurrent users, and the complexities of user queries (Gardner, 1998). A data warehouse scales both horizontally and vertically. The warehouse will grow as a function of data growth and the need to expand the warehouse to support new business functionality. Scalability can be defined using four dimensions (Fryer, 1998). The first, *environmental complexity*, involves supporting complex data models and queries. The second, *user concurrency*, refers to the number and types of queries that can be supported at the same time. Not all queries are of the same priority or complexity. The third dimension, *support for the environment*, must be planned as the data warehouse grows. The fourth dimension, *data volume*, refers to how will the current model and subsequent physical implementation support future growth. If the database is very large, it can affect the backup strategy. In many client server applications the use of RAID (redundant array of inexpensive disks) technology allows the database to be mirrored, so that in the event of a hardware problem, the mirrored copy may be used to keep the database operational. The warehouse design team will have to determine the cost benefit of implementing this approach for a potentially large data warehouse. In addition, the timing of the backups will have to be determined because the amount of data involved may require special processing.

DATA MART STRATEGIES

Some companies begin their movement towards a data warehouse strategy through the development of a data mart. A data mart can be either independent or dependent. An independent data mart is a smaller subject area data repository that is not directly connected to the enterprise data warehouse (Watson and Haley, 1998). An independent data mart is usually quicker to construct, and as such can serve as a proof of concept before a full-scale investment is made. This approach can also generate a quicker return on investment by realizing benefits sooner. This approach is viewed as an advantage by departments that are highly focused on control of their data. One of the major disadvantages in developing independent data marts is that with an enterprise data warehouse there is a single integrated data store, which reduces the potential for data integrity issues that arise when data are stored in multiple locations redundantly. An independent data mart strategy can lead to the proliferation of multiple independent silos of information, which can be a deterrent to developing a single integrated strategy. Further, it can prevent users from analyzing data using similar views that are made possible by the enterprise data warehouse.

A dependent data mart is a subset of the data warehouse organized by subject area (e.g. the marketing data mart). It provides the advantages of using a consistent data model and providing quality data. Dependent data marts support the concept of a single enterprise wide data model; however, they require that the data warehouse be constructed first.

For a data mart implementation to be successful, it is important to develop a scope up front and build it into the plan (Dyche, 1998). Scoping should determine how long the requirements definition phase should take. The scope should be narrowly focused on a solid business case. The scoping process should be a high-level evaluation designed to answer the significant questions that affect the development of the data mart. A common error in scoping is to combine the requirements gathering phase with the design phase. This can lead to prematurely modeling before the requirements are defined. One of the principle scoping questions, therefore, is who will translate the business requirements into the physical database design. There are several other issues to consider as well, such as the need for performance tuning, and how often will the data be refreshed. These myriads of issues highlights why data warehousing projects have typically been expensive, delivered behind schedule and require a significant amount of organizational resource.

Wixom and Watson (2001) defined a research model for data warehouse success that identified seven implementation factors. These factors can be categorized into three criteria (organizational issues, project issues, and technical issues). The factors consist of:

1. Management Support
2. Champion
3. Resources
4. User Participation
5. Team Skills
6. Source Systems
7. Development Technology

Further they point out those systems that display high data and system quality lead to increased net benefits. They provided an empirical measure to demonstrate the importance of user participation in the development of a data warehouse. Additionally, they highlighted the importance of the need to have strong management support for data warehouse projects.

Watson, Ariyachandra, and Matyska (2001) defined nine variables that help explain how organizations data warehouses change. These variables describe the different stages of maturity that data warehouses exhibit. The variables, data, architecture, stability of the production environment, warehouse staff, warehouse users, impact of users' skills, use of the warehouse, organizational impacts, and cost–benefits describe the factors that affect data warehouse usage.

TASK-TECHNOLOGY FIT

Goodhue and Thompson (1995) proposed that for information technology to have a positive impact on individual performance the technology must be utilized and it must be a good fit with the task that it supports. Task-technology fit provides a stronger theoretical basis for a number of issues related to the impact of information technology on individual performance, including understanding the impact of user involvement on performance. Performance impacts will result when a technology provides features and supports the fit of the requirements of the task. The higher the fit, the higher the performance increase will be. Goodhue and Thompson's research proposes that information systems impact performance only when there is a relationship between the task requirements of the user and the functionality of the system. The system must satisfy the business requirements of the user. Task-technology fit then is the degree to which a technology assists an individual in performing their tasks. Goodhue and Thompson (1995) developed eight measurement components of task-technology fit:

1. Data quality
2. Locatability of data
3. Authorization to access data
4. Data compatibility between systems
5. Training and ease of use
6. Production timeliness
7. System reliability
8. Information systems relationships with users

Performance impacts relate to the accomplishment of tasks by an individual. Improved efficiency, effectiveness or quality implies higher performance. High task-technology fit improves not only performance but also the likelihood of utilization, regardless of why the system is utilized. Utilization may be on a voluntary or mandatory basis. Goodhue and Thompson (1995) postulated that the characteristics of the task and the characteristics of the individual would affect user evaluations of task-technology fit. Task-technology fit (Goodhue, 1998) has been used to provide the basis for a user evaluation instrument aimed at an organizational assessment of information systems utilization for managerial decision-making. Measures of system usage have problems because it may not be clear if the utilization is a result of an effective system yielding greater efficiency or a poor system that requires greater effort to use. The heart of the task-technology fit model is the assumption that information systems that give value to users will be reflected in a user's evaluation of the systems (Goodhue, 1998). Information systems also need to change, as task needs change. Task characteristics will moderate the strength of the link between information system characteristics and user evaluations.

According to Goodhue (1998), in the analysis of tasks, information systems and services need to support users along the following dimensions:

1. Identification subtask – the right data, the data element definition, right level of detail.
2. Acquisition subtask – accessibility of information, ease of use of hardware and software, training, reliability, flexibility and cost.
3. Integration/interpretation subtask – compatibility, accuracy, presentation and currency.

Each dimension ascertains the degree to which users believe that task needs have been met by the information systems and services available to them. It will therefore be necessary to validate that each of these dimensions affects the utilization and performance of data warehouse environments. The task-technology fit model should lead to five propositions. Information systems and service characteristics, task characteristics, and individual characteristics should influence user evaluations of information systems. Task and individual characteristics should interact with or moderate the relationship between user evaluations and information systems. Higher task-technology fit should lead to higher individual performance (Goodhue, 1998).

Task complexity consists of three components (Zigurs and Buckland, 1998). Coordinate complexity is the number of non-linear sequences between components and task products. Component complexity is the number of distinct acts and the distinct information involved in the acts. Dynamic complexity is the stability of the relationships between inputs and the product. This allows complexity to be defined independent of the person performing the task. Complexity is critical to differentiate between distinct task environments. Campbell (1988) defined four dimensions of complexity level that are important in defining unique task environments. The first dimension, multiplicity, means there is more than one desired outcome for a task. An example of multiplicity is a task that has more than one stakeholder, each of whom have different expectations about the outcome of the task. The second dimension, solution scheme multiplicity, is defined as more than one possible course of action to attain a goal. This dimension of task increases information load. The third dimension of task complexity, conflicting interdependence, may exist among solution schemes where adopting one of them conflicts with adopting another possible solution scheme.

Conflict interdependence alters the decision process such that the decision-makers cannot simply undo that option and return to the same conditions in order to adopt a new choice. Conflict interdependence may also exist where information from one source is in conflict with other sources. The fourth dimension, solution scheme/outcome uncertainty is defined as the extent to which there is uncertainty about whether the solution is correct. When the outcome is not explicit, uncertainty is increased, such as when the scope of the problem is large and there is little historical information available. Combinations of the four dimensions of task complexity are determinants of sixteen different task environments. These task environments have been grouped into five task categories (Campbell, 1988). These task categories include: (1) Simple Tasks, (2) Problem Tasks, (3) Decision Tasks, (4) Judgment Tasks and (5) Fuzzy Tasks. Simple tasks consist of a single solution and outcome scheme.

Problem tasks are characterized by solution scheme multiplicity. They involve finding the best way to achieve the outcome, so that complexity is defined as the number of potential paths to the outcome. Decision tasks are characterized by outcome multiplicity. Decision task types are differentiated by either the presence or absence of conflicting interdependence among the multiple outcomes as well as by uncertainty. Decision tasks involve choosing a best alternative. Judgment tasks are characterized by conflicting interdependence. The person performing the task must first consider all sources of relevant information and then make a judgment or prediction on the likelihood of a future event. Fuzzy tasks are characterized by the presence of both outcome multiplicity and solution scheme multiplicity. Data warehouse environments are designed to support tasks of varying degrees of complexity simultaneously. It is expected that the dimensions of task complexity will have a significant effect on the user evaluation of a data warehouse.

PRIOR TASK-TECHNOLOGY FIT RESEARCH

Ferratt and Vlahos (1998) described the implications of task-technology fit on managerial decision-making. In their analysis they used four different views of managerial decision-making:

1. Anthony's view of decision making relative to the three traditional management functions, strategic management, management control and operational control.
2. Mintzberg's view of decision making in terms of the four decision roles of entrepreneur, disturbance handler, resource allocator, and negotiator.
3. Broad view of managerial decision making including identifying problems, determining alternative courses of actions, ranking the alternatives and choosing an outcome.
4. Isenberg's and Rockart and Delong's view that the manager uses a mental model of the organization as the basis for decision making and that a critical role of information systems is to support that mental model.

One of the results that Ferratt and Vlahos (1998) concluded was that the manager's evaluation of information systems was not related to the amount of time that they used the system. The amount of time that a system was used was one of their measures of fit. Its importance is in addressing the larger issue of understanding what types of systems organizations should invest in to support managerial decision-making.

Dishaw and Strong (1998) posited that software would be used if the functions available to the user support the activities of the user. Further, they proposed that a software function supports an activity if it facilitates the activity. They defined fit to be the matching of the capabilities of the technology to the demands of the task. Task-system fit is the degree that an information system or environment assists an individual in performing their portfolio of tasks. Their research focused on measuring task-technology fit in a software maintenance environment, which supported Goodhue and Thompson's (1995) assertion that a higher degree of fit lead to expectations of positive consequences of use by the individuals choosing to use the technology. They found that fit between maintenance tasks and maintenance tools is strongly associated with tool use. They developed a method for operationalizing and measuring task-technology fit using specific technology and task models to compute fit. Using the computed approach means that new questionnaire items to measure fit do not need to be developed each time the task-technology fit model is applied to specific task and technology environments. Further, Dishaw and Strong (1998) test the validity of task-technology fit in understanding the relationship between maintenance tasks and software tool usage. They extend the definition of task-technology fit to the following, "Task-technology fit is the matching of the functional capability of available software with the activity demands of the task" (Dishaw and Strong, 1999c). Their research concluded that tool experience is significant, has a positive effect on utilization, and is significant in the task-technology model.

Goodhue, Littlefield and Straub (1997) further extended the task-technology fit research by validating the model and using that validation to determine the impact of the Integrated Information Center on its target population. The validation of the model was achieved through the linkages that are used to create the task-technology fit model. Specifically, these include:

1. task-technology fit leads to utilization.
2. utilization and fit lead to performance impacts.
3. fit and performance impacts lead to feedback effects.

While they did not find the impact of the Integrated Information Center to be significant, they succeeded in producing a validated instrument to measure the task-technology fit model. Utilization, which is the percentage of the portfolio of tasks that an individual uses a technology to perform, was found to be an intervening variable between the individual performance and the task characteristics (Goodhue, 1997). Utilization can be measured by the degree to which technology utilization is institutionalized, by the number of different functions that are used, or by the duration of the usage. The performance impact of the technology will increase when there is high task-technology fit. Utilization will be high when the use of technology is important to the user, and when it gives them a greater competitive advantage in accomplishing tasks quickly, accurately, and with greater effectiveness.

Goodhue (1998) further extended the task-technology research to include user evaluations of information systems. He found empirical support for four propositions of task-technology fit that are determinants of user evaluations. User evaluations were found to be affected by the technology, the task and the individual. The value that is placed upon a technology by a user was dependent upon the task that was being performed. As the demands of the task increased, if the systems features met those demands then the users would respond more favorably. The user evaluations were found to be an accurate measure of the differences between the task and the system used to support the task. Users weighed the cognitive cost/benefit of the task-technology fit in their assessment of the subtasks that users performed. Important dimensions of task-technology fit that were evaluated were:

1. identifying the data that is needed.
2. accessing the data that is needed.
3. integrating and interpreting the accessed data.

This is significant in establishing a causal relationship between task-technology fit and utilization.

Mathieson and Keil (1998) explored the relationship between task-technology fit and the ease of use of the user interface of a system. They developed three hypotheses: (1) the task by itself does not determine perceived ease of use; (2) the information system by itself does not determine perceived ease of use; when the interface is controllable; and (3) the interaction of task and information system affects perceived ease of use. They concluded that neither the task nor the system explained performance. It was the combination of the task technology that was needed to explain performance. The subjects they tested perceived a greater ease of use when technology fit was high.

Task definition is essential in understanding the task-technology fit of group support systems. Zigurs and Buckland (1998) found that task definition is accomplished using the following: (1) Simple tasks should result in the best group performance when a configuration using communication support is used. (2) Problem tasks should result in the best group performance when using a configuration that supports information processing. (3) A configuration that supports process structuring is best in the performance of decision tasks. (4) A configuration that supports communication and information is best for judgment tasks. (5) A communication and information structure that also supports some process structuring is best for fuzzy tasks. Task-technology fit is particularly relevant to group support systems because it takes into account the attributes of tasks. Fit was linked to the criterion of group performance, which can be used in the larger context of studying the impact of technology on organizational performance. This is significant with data warehouse environments since there must be a fit for both individual and group usage.

Shirani, Tafti and Affisco (1999) measured task-technology fit for synchronous and asynchronous communication systems in order to determine its impact on group efficiency and organizational productivity. They focused on the use of differences in technology to determine if it impacted task fit. The degree of task structure was found to be the task characteristic that had the greatest impact. Structured tasks are routine with well-defined procedures in place to handle them; whereas unstructured tasks require higher degrees of group coordination due to the greater amount of inferential reasoning needed. Data warehouses are expected to support a large amount of unstructured tasks in order to support ad-hoc discovery.

Fry and Slocum (1984), Hrebiniak (1974), Randolph (1981) and David, Pearce and Randolph (1989) conceptualize three dimensions to define technology used by groups within an organization. The first dimension is task predictability, which is the degree to which a job is perceived to be familiar or unfamiliar. This will be a determinant of the number of exceptions that a group will face when using a system. The second dimension is problem analyzability, which is the extent to which a job becomes more difficult once an exception occurs. The degree to which technology is available to support the handling of exceptions will facilitate the generation of solutions or alternatives. The third dimension is interdependence that is the degree that individuals are dependent upon others in order to accomplish a task. The greater the interdependence, the greater the need to gather, analyze and distribute information; suggesting the greater the perceived usefulness of technology that supports information collection and dissemination. David, Pierce and Randolph (1989) suggest that group structure have a more significant impact on group performance than does technology. Further validation is required due to the extensive changes in technological capability since this study was performed. Additionally, however, they contend that the fit between technology and structure is a useful indicator of performance.

McCarthy, Aronson and Mazouz (2001) extended the concept of task-technology fit to show a significant relationship when applied to knowledge management systems. When the task to be performed was primarily unstructured, having the right information available was more significant than system availability and usage.

RESEARCH METHODOLOGY

Based upon the prior literature on task-technology fit, we wanted to test the applicability within a data warehousing environment. Data warehousing as a concept was developed in order to empower users to be able to extract and analyze information more efficiently and without the assistance of information technology professionals. Therefore, we expected to achieve a high task-technology fit when applied to this classification of technology end-users. Prior literature (Goodhue and Thompson, 1995) (Goodhue, 1995) has produced a validated survey instrument. We utilized that survey, modifying several of the questions to specifically pertain to data warehousing environments. The questionnaire (Appendix A) consisted of forty-four questions that utilized a five point Lickert scale, ranging from strongly disagree to strongly agree. The questions were designed to test the relationship between the individual and task characteristics and their affect on task-technology fit within data warehousing environments.

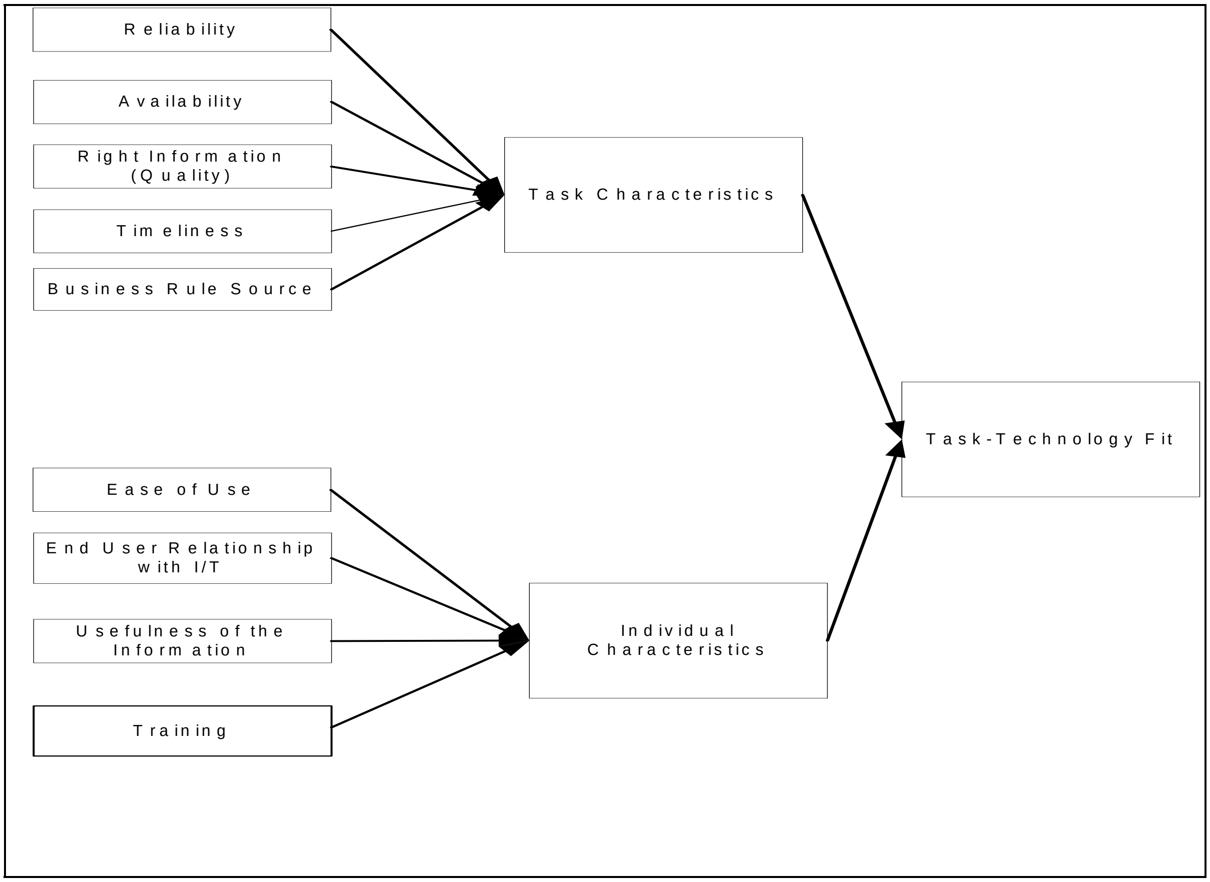
The survey was pre-tested using a group of fifty undergraduate and graduate students from Central Connecticut State University, a medium-sized university in the northeastern U.S. The graduate students were enrolled in a master's level class on database management systems and the undergraduate students were enrolled in a senior-level topics course specifically pertaining to data warehousing. They evaluated the survey for readability and understandability of the questions. Their results were not included in the analysis.

A total of one hundred and forty-four surveys were sent to data warehouse users within six companies. Sixty-six surveys were returned representing a response rate of 45.8%. The six companies were all within the insurance industry and were all domiciled in Connecticut. The surveys were distributed through a contact person at each of the companies, who was known to the researchers, which helps to explain the relatively high response rate.

The question that we were interested in testing is: What are the factors that affect task-technology fit in a data-warehousing environment?

Each of the questions within the survey was part of a construct for either a task characteristic or individual characteristic. The model that was adapted based upon the prior literature is described in Figure 1.

Figure 1: Data Warehousing Task-Technology Fit Model.



This research examines the relationship between the individual and task characteristics that impact the use of technology within a data warehouse environment to determine task-technology fit. Most of the independent variables to be tested are derived from the literature review and have been validated by prior task-technology fit research. The variables *business rule source* and *end user relationship with I/T* were developed based upon the data warehousing literature and the researcher's observations of three large-scale data warehouse projects. These variables consist of:

Reliability
Availability
Right Information (Quality)
Timeliness
Business Rule Source
Ease of Use
Training
Usefulness of the Knowledge
End User Relationship with I/T

These comprise task and individual characteristics. Task characteristics are affected by quality of the information, reliability, business rule source, availability, and timeliness. Business rule source refers to the origination point of the business rules that describe the information within the data warehouse. Business rule source is derived by either originating from the end user, the I/T organization or a combination of both. Ease of use, training, usefulness of the information, and the end user's relationship with the information technology data warehouse team affect individual characteristics. Individual characteristics combine to form the evaluation that a user makes as to whether a system provides value and will be used.

ANALYSIS OF THE RESULTS

Sixty-six survey responses were received; three contained missing values and were not considered as part of this analysis. Each of the questions applied to a specific construct. A factor analysis was then performed to identify correlated constructs. We used factor analysis to validate the constructs that comprised our variables. Two new variables are introduced as a result of this research; business rule source and end user relationship to the information technology organization. For each of the variables, factors whose coefficient was greater than 0.5 were used. This resulted in 8 questions being discarded. The results of the factor analysis are presented in Table 1.

Table 1: Factor Analysis of the Data Warehouse Variables.

Construct	Question Number	Coefficient
Quality (Q1)	Q3	0.5822
	Q11	0.6674
	Q26	0.6749
	Q33	0.6215
Usefulness (Q2)	Q6	0.6418
	Q23	0.5042
	Q29	0.5185
Reliability (Q13)	Q27	0.5215
	Q15	0.6060
Business Rule Source (Q7)	Q14	0.6196
	Q20	0.6771
	Q32	0.5875
	Q10	0.5817
End User Relationship with IT (Q9)	Q25	0.8100
	Q17	0.5186
	Q40	0.5860
Timeliness (Q18)	Q16	0.5003
	Q8	0.5691
Ease of Use (Q4)	Q21	0.5427
	Q28	0.6152
System Availability (Q5)	Q34	0.6704
	Q24	0.5808
	Q22	0.5047
	Q12	0.5112
Training (Q30)	Q35	0.5577
	Q37	0.5327
	Q39	0.6018

A Cronbach's Alpha test of construct validity was performed on each of the constructs that comprise the task and individual characteristics. It is generally held that a coefficient Alpha of 0.7 or greater indicates that the construct is reliable. The results (Table 2) indicated that the individual characteristics *end user relationship with I/T*, *training* and *ease of use* were shown to be highly reliable. However, the construct *usefulness (*0.6791) did not demonstrate a strong enough construct validity to be considered in further analysis and was therefore dropped. The task characteristics *quality*, *reliability, business rule source*, and *system availability* demonstrated strong construct validity. The construct *timeliness* was dropped from further analysis.

Table 2: Cronbach's Alpha – Task and Individual Characteristics.

Construct	Coefficient Alpha
Quality	0.8354
Reliability	0.7341
Business Rule Source	0.8084
Timeliness	0.6582
System Availability	0.7481
Usefulness	0.6791
End User Relationship with I/T	0.7871
Training	0.7132
Ease of Use	0.7072

End user relationship with I/T is a measure of the perceived cooperation between the users of the data warehouse and the system development team. Prior data warehousing literature has shown that to successfully build an enterprise data warehouse there needs to be a high degree of cooperation. Therefore, it is expected that there will be a higher level of task-technology fit in organizations that perceive the relationship between the end user and the I/T organization as stronger than those that do not.

Enterprise data warehouses rely on complex data models to capture, document and describe the idiosyncrasies of data relationships that can exist within a large organization. Data warehouse data models consist of the business rules that describe how information is defined and used within the organization. These business rules may come from a variety of sources. The end user might primarily define them or in some cases their definition is driven by the information technology organization. It is expected that the task-technology fit for data warehouses will be higher in organizations that do not primarily define their business rules via their information technology organization.

A Pearson's product moment correlation coefficient was calculated to determine the relationship of each of the variables to the task-technology fit. The results demonstrated strong correlations for both the task and individual characteristics at an $\alpha = .01$ (Table 3).

Table 3: Pearson's Correlation Coefficient.

Variable	Task-Technology Fit
Quality	0.266
Training	0.289
Reliability	0.389
Business Rule Source	0.360
System Availability	0.597
End User Relationship with I/T	0.571
Ease of Use	0.525

All of the variables under consideration for this study demonstrated a positive correlation. In particular, strong positive correlations exist for the impact of system availability, ease of use of a data warehouse and the end user's relationship with the information technology organization. Participation in the development and definition of the data model is expected to result in a greater acceptance and subsequent usage.

Multiple linear regression was then utilized to validate that there is a task-technology fit for data warehouse systems. Multiple linear regression has been used in prior studies (Goodhue and Thompson, 1995) (Dishaw and Strong, 1999c) (Dishaw and Strong, 1999a) to validate the relationship between individual and task characteristics with task-technology fit. The coefficient of determination (R^2) was calculated to be 0.647, at a confidence level of $\alpha = .01$. This suggests that the individual and task variables that were the basis for this study have a significant impact on task-technology fit within data warehousing environments.

DISCUSSION

Based upon the survey results, there is a demonstrated task-technology fit effect for data warehouse systems. A positive relationship was established between the individual and task characteristics that were the basis of the study. Further, we introduced two new variables that are applicable directly to data warehouse environments that help to explain task-technology fit when applied within this context. *Business rule source* and the *end user relationship with the IT organization* are directly related to task-technology fit. Both are key dimensions in the application development process and play a vital role when an enterprise data warehouse is architected. The size and complexity of an enterprise data warehouse necessitates that there should be a shared responsibility in the definition of business rules. This research is significant because it provides an empirical measure of the need to establish close ties between the end user and information technology units within organizations.

The theory of task-technology fit states that a system must meet the task requirements of a user in order for it to be utilized. End users must perceive that the system adds value, particularly when the purpose of the system (such as a data warehouse) is to improve decision-making. End user perception thus becomes a critical determinant of system acceptance.

Improving the decision making process within an organization is a much more subjective measure than operational functions due to the complexity in measuring results. End user perception thus becomes a critical determinant of system acceptance.

This study is significant because it provides a measure of the effectiveness of data warehouse systems. Data warehouse systems have been shown to be a very costly and resource intensive effort, that is frequently delivered late and over-budget. Expectations of the effectiveness of the data warehouse therefore begin to rise. If the system is not perceived as a value adding application, then it will have a negative impact on utilization. The results of this study suggest that one way to diminish that problem is to establish shared responsibility for business rule development when the system is being constructed.

This study is also significant because it demonstrated that timeliness and usefulness were not considered to be significant variables in the context of data warehouses. However, one of the limitations of this study is that the data warehouse users surveyed were all from within the insurance industry. Trend analysis in the insurance industry is usually measured over a period of years; therefore, the insignificance of timeliness of information may be attributable to the nature of the underlying business. Further study utilizing a cross-sectional approach across a variety of industries will validate the insignificance of timeliness as a variable. We propose that usefulness was not a significant variable in this study because the end user populations of data warehouse systems tend to be decision support specialists. The nature of their job tends to require a higher degree of understanding of the data and decision support tools that access the data, reducing the emphasis on usefulness.

The survey participants optionally had the opportunity to provide comments regarding their use of data warehouse system. Fourteen of the survey respondents provided comments. Ten of the responses centered on the critical need to have access to an integrated data dictionary. The respondents commented that it was essential to have access to detail information. Contact sensitive data dictionaries provide immediate feedback. The respondents commented that it was necessary to be able to reference documentation to more fully understand all of the information available within the data warehouse.

LIMITATIONS

One of the limitations of this study is that it only utilized data warehouse systems within the insurance industry. In particular, we feel this may have an impact on the results of the variable timeliness. The nature of the insurance business does not require immediate information to develop loss trends. Typically, information is gathered and evaluated over periods measured in years.

Our study also was limited by geographic region. All of the survey respondents were from the northeastern U.S. However, we feel they represented an accurate sample of data warehouse users.

FUTURE WORK

Two additional studies are planned as a result of this research. This study will be extended to determine if differences exist across industry segments. We intend to survey manufacturing, financial services, and retail services to determine if factor differences exist across industry segments.

A qualitative and quantitative study is planned to determine if there is a relationship between data modeling and end user acceptance of data warehousing systems. This study will attempt to analyze the data modeling factors that are significant in end user acceptance of data warehouse systems. The research will segment end users into two categories; one that represents data warehouse users that also participated in the development of the data model for the data warehouse. The second category will represent end users that did not participate in the development of the data model. A cross-sectional analysis will then be performed to determine if differences exist among the factors for these two groups.

APPENDIX A –
DATA WAREHOUSE USERS SURVEY

(Circle one per each line.)

		Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree
1	The data warehouse available to me is missing critical data that would be useful to me in my job.	1	2	3	4	5
2	The company maintains data at an appropriate level of detail for my purposes.	1	2	3	4	5
3	There are accuracy problems in the data I use or need.	1	2	3	4	5
4	It is easy to learn how to use the query tools that give me access to data.	1	2	3	4	5
5	The data model does not support my ad hoc reporting needs.	1	2	3	4	5
6	The data that is available to me allows me to make a significant contribution to the success of the business.	1	2	3	4	5
7	When business requirements change, it is easy to change the selection and format of data made available by your data warehouse.	1	2	3	4	5
8	It is easy to get assistance when I am having trouble finding or using the data.	1	2	3	4	5
9	It is easy to get access to data that I need.	1	2	3	4	5
10	On the reports I deal with, the exact meaning of data elements is available within the data dictionary or is easy to find.	1	2	3	4	5
11	Our data warehouse is too inflexible to be able to respond to my changing needs for data	1	2	3	4	5
12	I can count on the system to be “up” and available when I need it.	1	2	3	4	5
13	The data is up-to-date enough for my purposes.	1	2	3	4	5
14	Data that would be useful to me is unavailable because I don’t have the right authorization.	1	2	3	4	5
15	It is easy to find out what data the corporation maintains on a given subject.	1	2	3	4	5
16	I need some data on the up-to-the-minute status of operations or events but cannot get it.	1	2	3	4	5
17	Sometimes it is difficult or impossible to compare or aggregate data from two different sources because the data is defined differently.	1	2	3	4	5
18	I can’t get the data that is current enough to meet my needs.	1	2	3	4	5
19	It is easy to locate corporate or divisional data on a particular issue, even if I haven’t used that data before.	1	2	3	4	5
20	The data maintained by the corporation or division is exactly	1	2	3	4	5

		Strongly Disagree	Somewhat Disagree	Neither Agree Nor Disagree	Somewhat Agree	Strongly Agree
	what I need to carry out my tasks.					
21	The query tools that give me access to data are convenient and easy to use.	1	2	3	4	5
22	The data is subject to frequent system problems and crashes.	1	2	3	4	5
23	The data that I use or would like to use is accurate enough for my purposes.	1	2	3	5	5
24	I was involved in the development of the data warehouse.	1	2	3	4	5
25	The data is presented in a readable and useful format.	1	2	3	4	5
26	I am not getting as quick a turnaround from the data warehouse on requests for new reports or information.	1	2	3	4	5
27	The exact definition of data fields relating to my tasks is easy to find out.	1	2	3	4	5
28	I am getting the help I need in accessing and understanding the data.	1	2	3	4	5
29	The data that I need is displayed in a readable and understandable form.	1	2	3	4	5
30	It is hard to know how to use our query tool effectively.	1	2	3	4	5
31	It is more difficult to do my job effectively because some of the data I need is not available.	1	2	3	4	5
32	The corporation or division maintains sufficiently detailed understandable information.	1	2	3	4	5
33	There is not enough training on how to find, understand, access, or use corporate or divisional information.	1	2	3	4	5
34	Getting authorization to access information that would be useful in my job is time consuming and difficult.	1	2	3	4	5
35	I am getting the training I need to be able to use corporate or divisional data effectively in my job.	1	2	3	4	5
36	I can get information quickly when I need it.	1	2	3	4	5
37	When it's necessary to compare or aggregate information from two or more different sources, there may be unexpected or difficult inconsistencies.	1	2	3	4	5
38	The data warehouse was developed to meet the needs of a specific report(s).	1	2	3	4	5
39	I prefer to use the data warehouse query tool to others that I know of.	1	2	3	4	5
40	I have a good relationship with the IT professionals I deal	1	2	3	4	5

(Circle one per each line.)

		(Circle one per each line.)				
		Strongly Disagree	**Somewhat Disagree**	**Neither Agree Nor Disagree**	**Somewhat Agree**	**Strongly Agree**
	with.					
41	The data warehouse was originally designed for all of the ad hoc queries that I perform.	1	2	3	4	5
42	I rely on the data to make decisions.	1	2	3	4	5
43	The data is not up-to-date enough for my purposes.	1	2	3	4	5
44	The IT organization does not support my needs.	1	2	3	4	5

REFERENCES

Bajaj, A. and Nidumolu, S. (1998), A feedback model to understand information system usage. *Information & Management*, *33*(4), 213-224.

Campbell, D.J. (1988), Task complexity: a review and analysis. *Academy of Management Review*, *13*(1), 40-52.

Corey, M., Abbey, M., Abramson, I., Barnes, L., Taub, B., & Venkitachalam, R. (1999), *SQL Server 7 Data Warehousing*. Berkeley, CA: McGraw Hill.

David, F., Pearce, J. III, & Randolph, W. (1989), Linking technology and structure to enhance group performance. *Journal of Applied Psychology*, *74*(2), 233-241.

Davis, F., Bagozzi, R., & Warshaw, P. (1989), User acceptance of computer technology; a comparison of two systems. *Management Science*, *35*(8), 982-1001.

Dishaw, M., & Strong, D. (1998), Supporting software maintenance with software engineering tools: a computed task-technology fit analysis. *The Journal of Systems and Software*, *44* (February), 107-120.

Dishaw, M., & Strong, D. (1999a), Experience as a moderating variable in a task-technology fit model. *Proceedings of the Fifth Americas Conference on Information Systems*, Milwaukee, WI, Aug. 13-15, 1999.

Dishaw, M., & Strong, D. (1999b), Extending a model of task-technology fit to CASE tool utilization in design maintenance tasks. *Proceedings of the Fifth Americas Conference on Information Systems*, Milwaukee, WI, Aug. 13-15, 1999.

Dishaw, M., & Strong, D. (1999c), Extending the technology acceptance model with task-technology fit constructs. *Information & Management*, *36*, 9-21.

Dyche, J. (1998), Scoping your data mart implementation. *DBMS*, *11* (9), 43-50.

Ferratt, T.W., & Vlahos, G.E. (1998), An investigation of task-technology fit for managers in Greece and the US. *European Journal of Information Systems*, *7*, 123-136.

Fishbein, M., & Ajzen, I. (1975), *Belief, Attitude, Intention and Behavior: an Introduction to Theory and Research*, Reading, MA: Addison-Wesley.

Fry, L., & Slocum, J. (1984), Technology, structure, and work group effectiveness: A test of a contingency model. *Academy of Management Journal*, *27*, 221-246.

Fryer, R. (1998), Data-warehouse scalability. *Byte*, (June), 63-64.

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

Goodhue, D. (1997), The model underlying the measurement of the impacts of the IIC on the end-users. *Journal of the American Society for Information Science*, *48*(5), 449-453.

Goodhue, D. (1998), Development and measurement validity of a task-technology fit instrument for user evaluations of information systems. *Decision Sciences*, *29* (1), 105-138.

Goodhue, D. (1995), Understanding user evaluations of information systems. *Management Science*, *41*(12), 1827-1844.

Goodhue, D., Littlefield, R., & Straub, D. (1997), The measurement of the impacts of the IIC on the end users: the survey. *Journal of the American Society for Information Science*, 48(5), 454-465.

Goodhue, D., & Thompson, R. (1995), Task-technology fit and individual performance. *MIS Quarterly*, *19*(2), 213-236.

Hrebiniak, L. (1974), Job technology, supervision, and work group structure. *Administrative Science Quarterly*, *19*, 395-410.

Jarke, M., Jeusfeld, M., Quix, C., & Vassiliadis, P. (1999), Architecture and quality in data warehouses: an extended repository approach. *Information Systems*, *24* (3), 229-253.

Kimball, R. (1996). *The Data Warehouse Toolkit*, New York, NY: John Wiley & Sons.

Mathieson, K., & Keil, M. (1998), Beyond the interface: ease of use and task/technology fit. *Information & Management*, *34* (4), 221-230.

McCarthy, R., Aronson, J., & Mazouz, K. (2001), Measuring the validity of task-technology fit for knowledge management systems. *Proceedings of the Seventh Americas Conference on Information Systems*, Boston, MA, Aug 4-6.

Raden, N. (1997), Push back in push technology, *DBMS*, *10*, 63-68.

Randolph, W. (1981), Matching technology and the design of organizational units. *California Management Review*, *23*, 39-48.

Segars, A., & Grover, V. (1993), Re-examining perceived ease of use and usefulness: a confirmatory analysis. *MIS Quarterly*, *17*(4), 517-529.

Shirani, A., Tafti, M., & Affisco, J. (1999), Task and technology fit: a comparison of two technologies for synchronous and asynchronous group communication. *Information & Management*, *36*, 139-150.

Subramanian, A., Smith, L., Douglas, L., Nelson, A.., Campbell, J., & Bird, D. (1997), Strategic planning for data warehousing. *Information & Management*, (July) 99-113.

Watson, H., Ariyachandra, T., & Matyska, R. (2001), Data warehousing stages of growth. *Information Systems Management*, *18* (3), 42-50.

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

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

Zigurs, I., & Buckland, B. (1998), A theory of task/technology fit and group systems effectiveness. *MIS Quarterly*, *22* (3), 313-3.