December 2002

TASK-TECHNOLOGY FIT IN DATA WAREHOUSING ENVIRONMENTS: ANALYZING THE FACTORS THAT AFFECT UTILIZATION

Richard McCarthy
Quinnipiac University

Keith Lindsey
University of Memphis

Jay Aronson
The University of Georgia

Mark Frolick
University of Memphis

George Claffey
Central Connecticut State University

TASK-TECHNOLOGY FIT IN DATA WAREHOUSING ENVIRONMENTS: ANALYZING THE FACTORS THAT AFFECT UTILIZATION

Richard V. McCarthy
Quinnipiac University
richard.mccarthy@quinnipiac.edu

Jay E. Aronson
The University of Georgia
jaronson@terry.uga.edu

George Claffey
Central Connecticut State University
claffey@ccsu.edu

Abstract

Enterprise data warehouses can be 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?

The theory of task- technology fit defines a model that suggests that for a technology to be utilized it must meet the needs of a user. When the technology meets the users' needs and provides 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 is that there will be a high task-technology fit and that performance impacts will result in an improved organization.

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. Six organizations that have implemented a data warehouse were surveyed. The sample population consisted of users of large-scale data warehouses within the insurance industry.

Keywords: Data warehouse, task-technology fit

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, an organization must first develop an infrastructure to support 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 [23]. A data warehouse contains four characteristics [3] [30]:

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

Subject orientation enables users to determine not only how their business is performing, but why. A data warehouse differs from an operational database in that 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 on which each application type focuses [22]. Data warehouse *non-volatility* 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 [23]. 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 ([17], [20]) 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.

We present an overview of the development of the data warehouse as a critical component of an enterprise information systems architecture. We provide a brief theoretical foundation of task-technology fit theory as well as a description of 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 [28]:

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 [32]. 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 designing 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 [30]. A mapping approach should be used to provide a structured approach to classification of data. Data centric warehouses attempt to 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.

Task-Technology Fit

Goodhue and Thompson [20] 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 [20] developed eight measurement components of task-technology fit:

- Data quality
- Locatability of data
- Authorization to access data
- Data compatibility between systems
- Training and ease of use
- Production timeliness
- System reliability
- 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 [20] postulated that the characteristics of the task and the characteristics of the individual would affect user evaluations of task-technology fit. Task-technology fit [17] 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 [17]. 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.

Research Methodology

Based upon prior literature on task-technology fit, we wanted to test its 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 [18] [20] has produced a validated survey instrument. We utilized that survey, modifying several of the questions to specifically pertain to data warehousing environments. The questionnaire 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 in six companies. Sixty-six surveys were returned representing a response rate of 45.8%. The six companies were all in 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 shown in Figure 2.

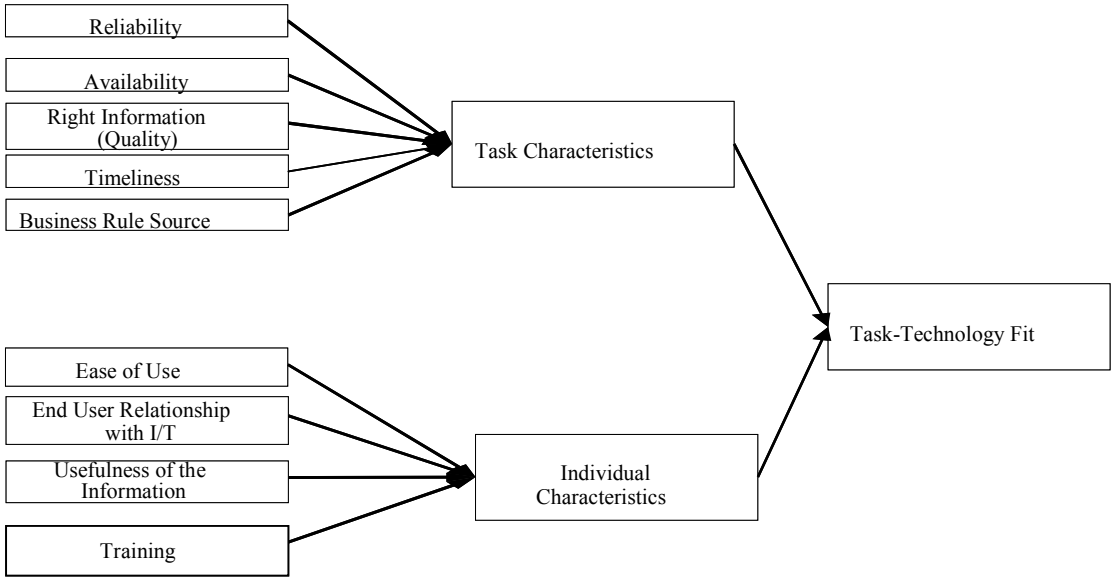

Figure 2. Data Warehousing Task-Technology Fit Model

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 researchers 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 confirmatory 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
	Q16	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. They might be primarily defined by the end user 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 α .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 [7] [9] [20] 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 designed. The size and complexity of an enterprise data warehouse necessitates that there 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 are 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 their 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.

References

Available upon request.