The Impact of Data Models and Task Complexity on End User Performance: An Experimental Investigation

By: Prashant Palvia and Chechen Liao

Abstract:

The purpose of this study was to investigate similarities and differences in the quality of data representations produced by end-users using the relational model (RM), the extended entity-relationship model (EERM), and the object-oriented model (OOM). By performing laboratory experiments using MIS major students, quality was evaluated on five constructs of a data model (i.e. entity/object, descriptor, identifier, relationship and generalization hierarchy) and six facets of a relationship (i.e. unary one-to-one, unary one-to-many, binary one-to-one, binary one-to-many, binary many-to-many and ternary many-to-many-to-many).

The research focused on two major issues: data model design and data model conversion. The first issue investigated the differences in user performance between the RM, the EERM and the OOM. The second investigated the differences in user performance between the RM and the relational conversions of the EERM and the OOM models. For the first issue, EERM and OOM scored much higher than the RM in correctness scores of binary one-to-many and binary many-to-many relationships, but only the EERM led to significance. The RM and OOM scored much higher than EERM for unary one-to-one relationships, however, only the RM resulted in significance. The OOM required significantly less time for task completion than EERM. For the second issue, RM and the relational conversion of OOM scored significantly higher than the relational conversion of EERM for unary one-to-one relationships.

Keywords: data models | end user performance | data representation | conceptual modeling

Article:

1. Background

End-user computing (EUC) has undergone explosive growth and received a great deal of attention among the MIS research community in recent years (e.g. Rockart & Flannery, 1983; Etezadi-Amoli & Farhoodmand, 1996; Speier & Brown, 1997; Nelson & Todd, 1999). EUC is commonly regarded as a significant and irreversible phenomenon in information systems development (e.g. Aggarwal, 1994; Yellen, 1997; Shayo, Guthrie & Igbaria, 1999). End users are mostly involved in environments in which database management systems (DBMSs) and fourth generation languages with DBMS capabilities are used as major tools for application development. The major effort of human factor research in database management systems has focused on issues related to query interfaces (e.g. Reisner, 1981; Jarke & Vassiliou, 1985). However, the advent of user-developed systems coupled with the innovation and proliferation of data models motivated us to study the usability of data models in this paper.

Data models are representation vehicles for conceptualizing user data requirements and design tools for facilitating the definition of data. The two widely known classes of data models which have been used or proposed for DBMS development are logical/implementation models and conceptual/semantic models. For convenience, the study will use the terms logical and conceptual to substitute for logical/implementation and conceptual/semantic, respectively. Among the three major logical models (i.e. hierarchical, network and relational), the relational approach is now extensively accepted and represents the dominant trend in marketplace (Date, 1990). Among conceptual models, entity-relationship model (Chen, 1976), semantic data model (Hammer & McLeod, 1981) and object-oriented model (e.g. Kim, 1990) have played major roles in research and/or practice. The relational model (RM), the entity-relationship model (ERM) and the object-oriented model (OOM) are included in this study. The semantic data model is not included as many of its concepts have been incorporated in the object-oriented model.

2. Human-computer interface model

According to the Hutchins, Hollan and Norman (1985) human-computer interface model, directness distance exists between a user's goals and knowledge of the application domain, and the level of description provided by the systems with which the user must deal. Directness refers to an impression or a feeling resulting from interaction with an interface while distance is used to describe factors which underlie the generation of the feeling of directness. The amount of user cognitive effort to manipulate and evaluate a system is directly proportional to this distance. Figure 1 is an adaptation of the Hutchins, Hollan, and Norman's human-computer interface model in the context of database design. There are two forms of distance: semantic and articulatory. Semantic distance reflects the relationship between the user intentions and the meaning of the data model. It is related to the distance between the semantics about real world and the meaning of constructs provided by the data model. Articulatory distance reflects the relationship between the physical form of the data model and its meaning.

This study attempts to test whether there is a significant difference among the relational model (RM), the extended entity-relationship model (EERM) and the object-oriented model (OOM) in semantic and articulatory distance. The EERM and OOM show relationships between entities/objects in a more explicit and direct fashion than RM. RM represents relationships in an indirect and implicit manner. Therefore, it is believed that EERM and OOM would facilitate less semantic and articulatory distance than RM. The OOM has been attempting to achieve even less

distance than the EERM. Coad and Yourdon (1990) stated that &&the primary motivation for identifying objects is to match the technical representation of a system more closely to the conceptual view of the real world" (p. 59). The objects attempt to model users' perceptions more closely than the (Kroenke, 1992).

IMPACT OF DATA MODELS AND TASK COMPLEXITY ON END-USER PERFORMANCE

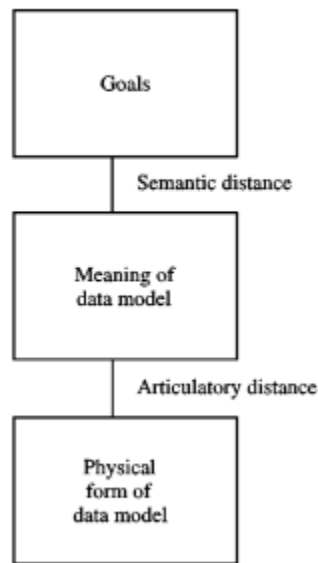

FIGURE 1. Semantic and articulatory distance in data modeling.

3. Previous research

Existing human factor studies in data modeling can be roughly divided into three categories. The first category is comparison among logical models, and typically focuses on the relational model vs. hierarchical and network models. For example, Brosey and Shneiderman (1978) found that the hierarchical model was significantly easier to use than the relational model, but only for the beginner group. Durding, Becker and Gould (1977) investigated how people organize data without using specific data models. Results suggested that the ease of use of a model is dependent on the inherent structure of data in an application, and the results supported the Brosey and Shneiderman findings.

The second category compares logical models with conceptual models, and has largely emphasized the relational model vs. conceptual models. Generally, the results favor one model or the other based on design task. Juhn and Naumann (1985) compared logical data structure (LDS), entity-relationship model (ERM), data access diagram (DAD) and relational model (RM). They reported that in relationship and cardinality finding tasks, ERM and LDS were superior to RM and DAD. On the other hand, RM outperformed ERM and LDS on identifier comprehension tasks. Ridjanovic (1986) found that subjects using LDS identified more relationships while subjects using RM identified more attributes. Jarvenpaa and Machesky (1989) found LDS superior to RM, especially in modeling entities and attributes. Batra, Hoffer, and Bostrom (1990) compared novice user performance using RM and EERM, and reported that EERM led to significantly better user performance in modeling binary and ternary relationships. Palvia (1991)

reported end-user's experience with hierarchical, network, relational and object-oriented models: OOM and network outperformed relational and hierarchical in terms of comprehension, efficiency and productivity. Liao and Shih (1998) investigated the effects of data models and training on data representation. Their results showed EERM to be superior to RM in many areas. Furthermore, the high degree training group outperformed the low degree one in modeling identifier, category and relationship.

4. Research questions

As discussed above, prior research addresses different logical and conceptual models in various combinations and permutations. Also, many of the findings provide mixed and conflicting results. At the present time, three data models clearly stand out: the relational model, the entity relationship model (or its extended version) and the object-oriented model. It is the purpose of this article to experimentally evaluate these three models (RM, EERM and OOM) on various dimensions from the perspective of the end-user. Furthermore, previous studies used only one task for the experiment and the characteristics of the single task itself may have favored a specific model. We bring in more rigor by including multiple tasks and multiple constructs. Accordingly, the following two main questions are addressed in this study.

1. What is the design effectiveness of the RM, EERM and OOM data models from end-users' perspective?

2. What is the quality of data representation of the relational design obtained directly from RM, and obtained after conversion from the EERM and OOM models?

The motivation for the second question arises from the practice of using EERM and the OOM models purely as conceptual models, and later converting them to a relational design prior to implementation.

5. Research methodology

5.1. RESEARCH MODEL

Jenkins (1982) factored the information system environment into four major elements: information system, human decision-maker, task and performance. Based on his conceptual model of the user-system interface, this study identifies four categories of variables which are vital to understanding database design and use*database management system (DBMS)/data model, human, task and performance. However, this study focuses only on the data model as our concern is the representation of data, and not data manipulation. A brief description for each variable in the research model (see Figure 2) follows.
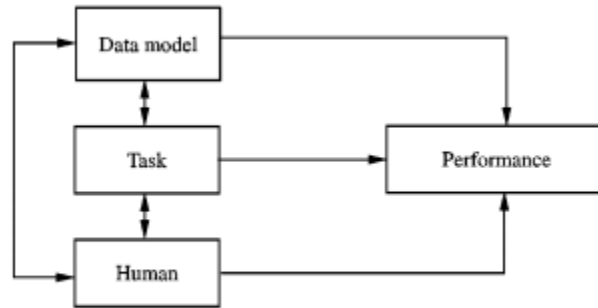
FIGURE 2. Research model for the study.

5.2. INDEPENDENT VARIABLES

5.2.1. Data model

Data models included in this study are relational model, extended entity-relationship model and object-oriented model.

Despite the popularity of relational implementations, the relational model has been criticized for its inability to capture certain semantics of the real world (Schmid & Swenson, 1975; Smith & Smith, 1977) and complex objects (Hammer & McLeod, 1981; Tsichritzis & Lochovsky, 1982), and its limitations that are inherent in record-based information modeling (Kent, 1979). The shortcomings and limitations of the relational model have led to proposals for conceptual models (e.g. Chen, 1976; Smith & Smith, 1977; Hammer & McLeod, 1981; Hull & King, 1987; Kim, 1990; Kroenke, 1992).

Entity-relationship model (Chen, 1976) along with its subsequent extended version (Elmasri, Weeldreyer & Hevner 1985; Teorey, Yang & Fry, 1986) defines an application as a set of identifiable entities, relationships between entities and their associated attributes using a graphical representation techniques. The entity-relationship representation can be converted to a relational representation for database implementation. This study adopts the extended version of ERM (EERM) as one of the three representation tools in the experiment.

This study adopts Kim's (1990) core object-oriented model concepts as a blueprint combined with Kroenke's (1992) semantic object model which is capable of designing semantic objects and converting the objects into relational representation for database implementation. The core modeling concepts in Kim's (1990) core object-oriented model include object and object identifier, attributes and methods, encapsulation and message passing, class and class hierarchy and inheritance. However, this research only deals with the representation of data, not data manipulation. Therefore, the methods, encapsulation and message passing are not included.

TABLE 1
A comparison of two tasks with different levels of complexity

Item	Task 1 (low complexity)	Task 2 (high complexity)
Entity/object	6 entities/objects	11 entities/objects
Descriptor	8 descriptors	15 descriptors
Identifier	6 identifiers (for EERM and OOM) 7 identifiers (for RM)	11 identifiers (for EERM and OOM) 14 identifiers (for RM)
Relationship	1 one-to-one unary	1 one-to-one unary
		1 one-to-many unary
	1 one-to-one binary	1 one-to-one binary
	1 one-to-many binary	1 one-to-many binary
	1 many-to-many binary	2 many-to-many binary
		1 many-to-many-to-many ternary
Generalization hierarchy	1 generalization hierarchy	2 generalization hierarchies

5.2.2. Task factor

Task characteristics are peculiar to the problem domain. Task structure and task complexity may affect the user's performance. In this study, two levels of complexity are included. Task 1 and 2 represent low and high levels of complexity, respectively. The complexity is based on the numbers of entities/objects, the degree of relationships between entities/objects, and the degree of nesting of entities/objects, relationships and generalization hierarchies. The tasks are presented as narratives and the subjects are asked to develop a data model using one of the three data models discussed earlier. A comparison of the two tasks is presented in Table 1.

5.2.3. Control variable

Human characteristics, such as programming experience, level of computing skills, database experience, data modeling experience, work experience, age and education may interact with the data models and have significant effects on user performance. In this study, the human factor serves as a control variable. A class of end-users with relatively uniform degree of training and experience participated in the experiments. These users possessed a moderate amount of computing skills to develop and use their own applications.

5.2.4. Dependent variables

Modeling correctness is the primary variable for user performance measurement. Modeling correctness is defined as the degree to which a data representation approaches the correct solution, whereas the correct solution conveys the same semantics about data as the natural language description of the database application (Batra et al., 1990). Modeling correctness will be measured through five different constructs of the data model: entities/objects, descriptors, identifiers, relationships and generalization hierarchies, and six facets of a relationship: unary one-to-one relationship, unary one-to-many relationship, binary one-to-one relationship, binary one-to-many relationship, binary many-to-many relationship and ternary many-to-many-to-many relationship.

Efficiency is also used to evaluate performance. Efficiency is defined as the time required by end-users to complete the task satisfactorily. Based on prior relevant studies (Batra et al., 1990), the study defined &&satisfactorily" as when end-users complete the task by achieving an average percentage score no less than 60%. Thus, for the efficiency variable, only those users with a score of 60 or more are included in the analysis.

Perceived ease of use is also selected as a dependent variable. Davis (1989) defined perceived ease of use as the degree to which an individual believes that using a particular system would be free of physical and mental effort. The perceived ease-of-use instrument was adapted from Batra et al. (1990). The study added one more question to the instrument asking the subjects to express overall confidence in the solution they prepared.

6. Hypotheses

Specific hypotheses derived from the two research questions are stated in null forms. Hypotheses 1-24 are derived from research question 1. Hypotheses 25-42 are derived from research question 2. Table 2 shows the relationship between hypotheses from questions 1 and 2 and dependent variables.

Hypotheses 1-8 deal with the main effects of the independent variable, data model. They are used to investigate the difference between RM, EERM and OOM in semantic and articulatory distance in terms of user performance. An example (H1) of these hypotheses is worded as: there will be no significant difference in overall user performance between RM, EERM and OOM in the modeling of entities/objects.

Hypotheses 9}16 deal with the main effects of the independent variable, task. They are used to investigate the difference in user performance between tasks 1 and 2. An example (H9) of these hypotheses is worded as: there will be no significant difference in overall user performance between tasks 1 and 2 in the modeling of entities/objects.

Hypotheses 17-24 deal with the interaction between the three data models and the two tasks. An example (H17) of these hypotheses is worded as: there will be no significant difference between RM, EERM and OOM in user performance over tasks 1 and 2 in the modeling of entities/objects.

Hypotheses 25-30 are used to investigate the difference between the quality of relational representation directly using RM and the quality of relational representation converted from EERM and OOM. The quality here refers to the level of modeling correctness. An example (H25) of these hypotheses is worded as: there will be no significant difference in overall users1 performance between the relational representation directly using RM and the relational representation converted from EERM and OOM in the modeling of entities/objects.

Hypotheses 31-36 are used to investigate the difference between tasks 1 and 2 in the quality of relational representation obtained from the three data models. An example (H31) of these hypotheses is worded as: there will be no significant difference between tasks 1 and 2 in overall users1 performance in modeling the relational representation of entities/objects.

TABLE 2
Relationship between hypotheses from questions 1 and 2 and dependent variables

Hypotheses			Dependent variable
Main effect		Interaction	
Model	Task		
Research question 1 (data model design)			
H1	H9	H17	Entities/objects
H2	H10	H18	Descriptors
H3	H11	H19	Identifiers
H4	H12	H20	Relationships
H4a	H12a	H20a	Unary one-to-one relationship
H4b	H12b	H20b	Unary one-to-many relationship
H4c	H12c	H20c	Binary one-to-one relationship
H4d	H12d	H20d	Binary one-to-many relationship
H4e	H12e	H20e	Binary many-to-many relationship
H4f	H12f	H20f	Ternary many-to-many-to-many relationship
H5	H13	H21	Generalization hierarchies
H6	H14	H22	Efficiency
H7	H15	H23	Perceived ease-of-use
H8	H16	H24	Confidence level
Research question 2 (data model conversion)			
H25	H31	H37	Entities/objects
H26	H32	H38	Descriptors
H27	H33	H39	Identifiers
H28	H34	H40	Relationships
H28a	H34a	H40a	Unary one-to-one relationship
H28b	H34b	H40b	Unary one-to-many relationship
H28c	H34c	H40c	Binary one-to-one relationship
H28d	H34d	H40d	Binary one-to-many relationship
H28e	H34e	H40e	Binary many-to-many relationship
H28f	H34f	H40f	Ternary many-to-many-to-many relationship
H29	H35	H41	Generalization hierarchies
H30	H36	H42	Efficiency

Hypotheses 37-42 deal with the interaction between the three data models and the two tasks. An example (H37) of these hypotheses is worded as: there will be no significant difference between the relational representation directly using RM and the relational representation converted from EERM and OOM in user performance over tasks 1 and 2 in the modeling of entities/objects.

7. Research strategy

Several types of research materials were developed to conduct the investigation. These included: (a) a questionnaire for demographics and computer experience data, (b) a set of training notes, (c) two textual cases (tasks 1 and 2) describing organizational data requirements, (d) a questionnaire for perceived ease of use and confidence level, (e) solutions for the organizational database applications, (f) itemized solutions for tasks 1 and 2 and (g) a grading scheme.

Two months before the actual experiment, a pilot study was conducted to identify procedural problems, validate the research instrument and collect other useful information. Eighteen volunteer MBA students participated in the study. The pilot provided useful information about

procedural problems, the time required for training and task completion and the subjects' ability in preparing data models for non-trivial database applications. Several changes were made in the final procedures including the grading scheme.

After the pilot, laboratory experiments were conducted and actual measurements were made. The 66 subjects were students in junior and senior classes of the MIS program in an American University. The experiments were conducted during the normal class schedule. Data model training was conducted separately in three classes. The data model application was then given to these classes one week later. Before the training, subjects were told that they were required to participate in both data model training and data model application. Subjects were also informed that the data model application would serve as a test and would be graded. They would receive credit for the test as part of their final course grade. This helped to ensure a higher level of motivation.

The actual experiment included the following steps.

1. The subjects were asked to complete a questionnaire regarding personal demographics and computer experience.

2. The subjects were then provided with a set of notes and were trained by the experimenter in using one of the data models for database design. The subjects were informed that they could use the notes for data model design. The training lasted 55 min for the RM group, and 75 min each for the EERM and OOM groups. The extra 20 min for the EERM and OOM groups were used to tell subjects how to convert the EERM and OOM to RM. Note that this amount of training is comparable to similar studies in the past (e.g. Batra et al., 1990). Besides, end-users typically tend to have very little training in such tasks.

3. The subjects were then provided with the case description and an answer sheet, and were asked to design the database using the assigned data model.

4. After finishing the data model design, each subject was asked to complete a questionnaire regarding perceived ease-of-use and user overall confidence.

5. The subjects for the EERM and OOM groups were then provided with another answer sheet. In this, they were required to convert the EERM/OOM to RM.

8. Analysis and results

8.1. SUBJECT DEMOGRAPHICS

Subjects consisted of 66 MIS major undergraduate students: 57 seniors and 9 juniors. They are from two sections of a system analysis course and one section of an IS planning course. Since the experiment was conducted in a normal class schedule, randomly assigning subjects to the three treatment groups was not possible. However, each class was randomly assigned one data model. In each class, subjects were randomly assigned either task 1 or task 2. A cross-tabulation procedure of SPSS was used to examine the subjects' characteristics between classes. The chi-

square likelihood ratio test was used. Results indicated that no significance was found in the treatment groups in terms of the subjects' characteristics.

8.2. INTERACTION BETWEEN DATA MODEL AND TASK

Hypotheses 17-24 (research question 1) and hypotheses 37-42 (research question 2) deal with the interaction effects between data model and task. The first group of hypotheses investigated the difference in user performance among RM, EERM and OOM. The second group investigated the difference in user performance among RM, relational conversion after EERM and relational conversion after OOM. Table 3 presents results for the two-factor analysis of variance. As seen, no significance was found in any of these hypotheses. This means that there is no evidence to suggest that a task with low complexity may favor a specific data model while a task with high complexity may favor other data models.

8.3. DATA MODEL

8.3.1. Relationship Hypotheses

1-8 (research question 1) and hypotheses 25-30 (research question 2) deal with the main effects of data model. The first group of hypotheses investigated the difference in user performance among RM, EERM and OOM. This part was related to data model design. The second group investigated the difference in user performance among RM, EERM relational conversion and OOM relational conversion. Table 3 presents results for the two-factor analysis of variance. As seen, significant differences were found in the unary one-to-one relationship, binary one-to-many relationship and binary many-to-many relationship in the design part. Significant difference was found only in the unary one-to-one relationship in the conversion part. Table 4 compares results of modeling constructs between design and conversion. An interesting observation is that there was a sharp drop in the mean scores of the binary one-to-many relationship and the binary many-to-many relationship after the EERM and OOM were converted to the relational representation. On the other hand, there was only a very slight change in the mean scores of the unary one-to-one relationship when these two models were converted to relational forms. Therefore, the significant differences found in the design part for the binary one-to-many and the binary many-to-many relationship disappeared in the conversion part. The unary one-to-one relationship, however, retained significance in both design and conversion parts.

In design, subjects using RM performed significantly better than those using EERM in modeling the unary one-to-one relationship. The OOM group also scored more than the EERM group by 17.27%, although this was not statistically significant. In conversion, both RM and OOM groups outperformed the EERM group in modeling the unary relationship. These results can be possibly attributed to the fact that RM and OOM both provide a more direct and simple way of modeling the unary relationship than does EERM. In EERM, a unary relationship is captured by a relationship symbol connected to the same entity, which is a somewhat difficult concept.

TABLE 3
Summary of results for questions 1 and 2

Dependent variable	Interaction		Main effect model		Main effect task	
	Hypo	Sig.	Hypo	Sig.	Hypo	Sig.
Research question 1 (data model design)						
Entities/objects	H17	NS	H1	NS	H9	NS
Descriptors	H18	NS	H2	NS	H10	NS
Identifiers	H19	NS	H3	S*	H11	NS
Relationships	H20	NS	H4	NS	H12	NS
Unary one-one	H20a	NS	H4a	S***	H12a	NS
Unary one-many	H20b	N/A	H4b	NS	H12b	N/A
Binary one-one	H20c	NS	H4c	NS	H12c	NS
Binary one-many	H20d	NS	H4d	S**	H12d	NS
Binary many-many	H20e	NS	H4e	S***	H12e	NS
Ternary many-many-many	H20f	N/A	H4f	NS	H12f	N/A
Generalization hierarchies	H21	NS	H5	NS	H13	NS
Efficiency	H22	NS	H6	S**	H14	S***
Perceived ease-of-use	H23	NS	H7	NS	H15	NS
Confidence level	H24	NS	H8	NS	H16	NS
Research question 2 (data model conversion)						
Entities/objects	H37	NS	H25	NS	H31	NS
Descriptors	H38	NS	H26	NS	H32	NS
Identifiers	H39	NS	H27	NS	H33	NS
Relationships	H40	NS	H28	NS	H34	NS
Unary one-one	H40a	NS	H28a	S***	H34a	NS
Unary one-many	H40b	N/A	H28b	NS	H34b	N/A
Binary one-one	H40c	NS	H28c	NS	H34c	S*
Binary one-many	H40d	NS	H28d	NS	H34d	NS
Binary many-many	H40e	NS	H28e	NS	H34e	NS
Ternary many-many-many	H40f	N/A	H28f	NS	H34f	N/A
Generalization hierarchies	H41	NS	H29	NS	H35	NS
Efficiency	H42	NS	H30	NS	H36	NS

*Significant at a 0.10 level.
**Significant at a 0.05 level.
***Significant at a 0.01 level.

However, the RM group scored significantly less than the EERM group for binary one-to-many relationships (by 29.58%) and for binary many-to-many relationships (by 33.23%). The RM groups also scored less than the OOM group for binary one-to-many relationships (by 19.69%) and for binary many-to-many relationship (by 24.19%), although these differences were not statistically significant. A plausible explanation is the following. The relational model represents a binary one-to-many relationship by placing the identifier of the parent relation in the child relation, while it represents a binary many-to-many relationship by creating a third relation (called intersection relation). This is somewhat artificial, complicated and an inconsistent manner, at least to the naive end-user. The problem is exacerbated when the EERM and OOM groups convert their EERM and OOM designs to relational forms. The degree of drop in the means of the binary many-to-many relationship was higher than that in the binary one-to-many relationship for both the EERM and OOM groups since the subjects had difficulty in creating the intersection relation

As seen in Table 4, no significant differences were found in both design and conversion parts for the other relationship variables. However, there was also a sharp decline in the mean scores of the relationship variables when the EERM and OOM were converted to relationship forms. The dramatic drop implies that the subjects had difficulty in converting the EERM/OOM to the relational representation.

8.3.2. Identifiers

At p=0.07 (slightly higher than the usual significance level of 0.05), there was a difference in the mean scores of identifiers between the three data models in the design part, while no significance was found in the conversion part. In the design part, the Tukey follow-up test showed that the OOM group performed significantly better than the RM group. The OOM facilitates a clear and direct method to model identifiers, and this resulted in its superiority. On the other hand, in the RM, identifiers are also used to define relationships. In effect, RM provides an implicit and indirect way to model relationships. The RM group's poor performance in modeling the relationships affected their performance in modeling identifiers. Although the EERM and OOM groups scored more than the RM group, after conversion of the EERM and OOM to relational forms, the higher mean scores were sharply reduced.

8.3.3. Efficiency

In the design phase, there was a significant difference (p value"0.031) in the means of user efficiency for task completion among the three data models. The mean scores for RM, EERM and OOM were 37.67 42.38, and 32.39 min, respectively. The Tukey follow-up test indicated that the OOM group required significantly less time for task completion than the EERM group. The EERM group required more time for task completion possibly because of the more complex notation in EERM. However, there was no significant difference (p value"0.455) in the means of users' efficiency for model conversion between the EERM and OOM groups. The means for EERM and OOM are 15.80 and 17.38, respectively. Since the RM group did not have to do the conversion, the RM was not included in the analysis.

8.4. TASK FACTOR

Subjects in task 1 scored more than subjects in task 2 in most of the dependent variables. However, a significant distinction in user performance between tasks 1 and 2 was not obtained for the overall tasks. Actually, the effect of task complexity is observed in the task's individual components as reported earlier. The task complexity manifests itself through a number of characteristics, such as the numbers and nesting of entities/objects, relationships generalization hierarchies and the degree of relationships. Thus, while the task size effect could not be directly observed, perhaps due to the limits on our experiments, task complexity's effect was evidenced in the modeling of the different constructs.

9. Implications and further research

The major differences among the data models were from the relationship constructs. For the unary one-to-one relationship, this study reveals that RM and OOM can capture more semantics

than does EERM since they provide a more direct and simple method. Despite the superiority of RM in modeling the unary one-to-one relationship, the relational model represents the binary relationship in a more implicit and indirect manner than the other two models. As a consequence, the results indicate that the relational model is inferior to the other two models. The binary relationship occurs frequently in real-world applications, and this mainly contributes to the problems with the relational representation. However, while EERM is generally superior in representing relationships, it does require significantly more time to construct than RM and OOM. The more complex notation in EERM suggests than end-users may actually require more training on it.

Although the EERM and OOM groups scored more than the RM group in modeling relationships, after conversion to relational representation, the RM group scored more than the EERM and OOM groups. This implies that the EERM and OOM groups suffered when they used the relational representation to do the conversion. The advantage gained by the EERM/OOM was subsequently lost. Therefore, after the conceptual data model is ready, the end-users may need help from a professional (e.g. a DBA). Another solution is that a software package for database design may be used. End-users may use the package as an aid in the design of the conceptual data model. Once completed, the package will automatically convert the conceptual data model to the relational form.

This study also points to the types of weakness and errors that occur in each model. This knowledge can serve as a basis for facilitating a better understanding of end-users' capability of designing data models. Necessary training and support could be provided to improve end-user performance. The relationship construct probably requires the maximum training and support since most errors pertain to it.

Several extensions to this study are possible. First, a field setting can be conducted to validate the findings of this study. In a field setting, actual data modeling applications may be developed and compared with the results of our study. Second, the human factor served as a control variable in this study. The extension to include this factor as an independent variable will be useful to understand the effect of expertise on the data modeling task. For instance, a future study could include both end-users and expert designers. Third, another extension would be to include other object-oriented models with greater functionality. Finally, future research can be also extended to consider specific human characteristics on data modeling performance, such as cognitive style.

References

AGGARWAL, A. K. (1994). Trends in end user computing: a professional's perspective. Journal of End User Computing, 6, 32}33.

BATRA, D., HOFFER, J. A. & BOSTROM, R. P. (1990). Comparing representations with relational and EER models. Communications of the ACM, 33, 126-139.

BATRA, D. & KIRS, P. J. (1993). The quality of data representations developed by non-expert designers: An experimental study. Journal of Database Management, 4, 17-29.

BOCK, D. B. & RYAN, T. (1993). Accuracy in modeling with extended entity relationship and object oriented data models. Journal of Database Management, 4, 30-39.

BROSEY, M. & SHNEIDERMAN, B. (1978). Two experimental comparisons of relational and hierarchical database models. International Journal of Man-Machine Studies, 10, 625-637.

CHEN, P. P. (1976). The entity-relationship model – toward a unified view of data. ACM Transactions on Database Systems, 1, 9-36.

COAD, P. & YOURDON, E. (1990). Object-Oriented Analysis. Englewood Cliffs, NJ: Yourdon Press.

DATE, C. J. (1990). An Introduction to Database Systems (5th edn), Vol. 1. Reading, MA: Addison-Wesley.

DAVIS, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13, 319-340.

DURDING, B. M., BECKER, C. A. & GOULD, J. D. (1977). Data organization. Human Factors, 19, 1-14.

ELMASRI, R., WEELDREYER, J. & HEVNER, A. (1985). The category concept: an extension to the entity-relationship model. Data & Knowledge Engineering, 1, 75-116.

ETEZADI-AMOLI, J. & FARHOODMAND, A. F. (1996). A structural model of end user computing satisfaction and user performance. Information and Management, 30, 65}73.

HAMMER, M. & MCLEOD, D. (1981). Database description with SDM: a semantic database model. ACM ¹ransactions on Database Systems, 6, 351}386.

HULL, R. & KING, R. (1987). Semantic database modeling: survey, applications and research issues. ACM Computing Surveys, 19, 201}260.

HUTCHINS, E. L., HOLLAN, J. D. & NORMAN, D. A. (1985). Direct manipulation interfaces. Human-Computer Interaction, 1, 311}338.

JARKE, M. & VASSILIOU, Y. (1985). A framework for choosing a database query language. Computing Surveys, 17, 313}340.

JARVENPAA, S. L. & MACHESKY, J. J. (1989). Data analysis and learning: an experimental study of data modeling tools. International Journal of Man-Machine Studies, 31, 367}391.

JENKINS, A. M. (1982). MIS Decision Variable and Decision Making Performance. Ann Arbor, MI: UMI Research Press.

JUHN, S. & NAUMANN, J. D. (1985). The effectiveness of data representation characteristics on user validation. Proceedings of the 6th International Conference on Information Systems, pp. 212-226.

KENT, W. (1979). Limitations of the record based information models. ACM Transactions of Database Systems, 4, 107-131.

KIM, W. (1990). Introduction to Object-Oriented Databases. Cambridge, MA: The MIT Press.

KROENKE, D. M. (1992). Database Processing (4th edn.). New York: Macmillan Publishing Company.

LIAO, C. & SHIH, M. (1998). The effects of data models and training degrees on end users' data representations. MIS Review, 8, 1-20.

LIAO, C. & WANG, L. (1997). The comparison of EERM and OOM on novice user performance. Information Management Research, 1, 25-48.

NELSON, R. R. & TODD, P. (1999). Strategies for managing EUC on the Web. Journal of End User Computing, 11, 24-31.

PALVIA, P. (1991). On end-user computing productivity. Information and Management, 21, 217-224.

REISNER, P. (1981). Human factor studies of database query languages. ACM Computing Surveys, 13, 13-31.

RIDJANOVIC, D. (1986). Comparing quality of data representations produced by non-experts using logical data structure and relational data models. Unpublished Ph.D. Dissertation, University of Minnesota.

ROCKART, T. F. & FLANNERY, L. S. (1983). The management of end-user computing. Communications of the ACM, 26, 776-784.

SCHMID, H. A. & SWENSON, J. R. (1975). On the semantics of the relational data model. Proceedings of ACM SIGMOD, pp. 211-223.

SHAYO, C., GUTHRIE, R. & IGBARIA, M. (1999). Exploring the measurement of end user computing success. Journal of End User Computing, 11, 5-14.

SHOVAL, P. & SHIRAN, S. (1997). Entity-relationship and object-oriented data modeling – an experimental comparison of design quality. Data and Knowledge Engineering, 21, 297-315.

SPEIER, C. & BROWN, C. V. (1997). Differences in end-user computing support and control across user departments. Information and Management, 32, 85-99.

TEOREY, T., YANG, D. & FRY, J. P. (1986). A logical design methodology for relational databases using the extended entity-relationship model. ACM Computing Surveys, 18, 197-222.

TSICHRITZIS, D. C. & LOCHOVSKY, F. H. (1982). Data Models. Englewood Cliffs, NJ: Prentice-Hall.

YELLEN, R. E. (1997). End user computing in a global environment. Journal of End User Computing, 9, 33-34.