An Experimental Comparison Among Conceptual Data Modeling Techniques

이희석
한국과학기술원 테크노경영대학원 부교수
최병구
한국과학기술원 테크노경영대학원 박사과정

Abstract

This paper compares conceptual data models including extended entity relationship (EER), semantic object model (SOM), object role modeling (ORM), and object modeling technique (OMT) in terms of model correctness, time to understand, and perceived ease-of-use. For an empirical study, 28 graduates and 72 undergraduates were selected and then divided into four equally sized groups. Each group was trained with one data modeling technique. Two cases were used; one was prepared in natural language and the other in enterprise form. The study results show some differences among the four conceptual data modeling techniques. These positive findings may help modelers better understand modeling techniques.

I. Introduction

By using a conceptual data model one can describe a reality. Conceptual data modeling technique makes it easy to understand and interpret reality; so, selecting conceptual data modeling technique is very important for effective database design. The quality of the resulting schema depends not only on the skill of the database designers and users, but also on the qualities of the technique selected (Batini et al., 1992). Advocates of each conceptual data modeling technique assert that his or hers is better in model correctness and shorter time required in solving problem, but, there is little evidence to support their assertions (Kim and March, 1995).

This paper compares four conceptual data modeling techniques: EER (Extended-Entity-Relationship), SOM (Semantic Object Model), ORM (Object Role Modeling), and OMT (Object Modeling Technique). EER is an enhanced version of an Entity-Relationship (ER) model, the most popular data modeling technique that was proposed by Chen (1976). SOM was originally developed by Kroenke (1995); it is very similar to EER, but does not include the concept of relationship; all relationships are represented in terms of attributes. Halpin (1995) proposed ORM, based on NIAM (Nijssen Information Analysis Methodology) that was developed in Europe in the 1970s (Nijssen, 1994). It is widely used in Australia and Europe. OMT is an improvement of ER, and Logical Relational Design Methodology (LRDM) model (Blaha et al., 1988 ; Teorey et al., 1986).

For model comparison, Batra et al. (1990) used model correctness and perceived ease-of-use. Bock and Ryan (1993) used model correctness. Hardgrave and Dalal (1995) used model understanding, understanding time, and perceived ease-of-use. In this paper, we compare conceptual data modeling techniques in terms of model correctness, understanding time, and perceived ease-of-use. Model correctness can be defined as the degree to which a conceptual model approaches the correct solution(s). Understanding time is time duration needed for the modeler to understand the model (Shoval and Even-Chaime, 1987; Hardgrave and Dalal, 1995). For simplicity, time is measured by the time needed for modeling each case. Perceived ease-of-use is defined as the degree to which a person believes that using a particular technique would be free effort. It is measured by the use of questionnaires developed by Davis (1989).

In addition, we perform two other studies on data modeling techniques. First, we study the interaction effect between data modeling techniques and past experiences on EER; second, we study the correlation between natural language case and enterprise form case.

II. Literature Review

Prior studies can be classified into three categories: (i) comparison of classical data models, (ii) comparison of conceptual data model with classical model, and (iii) comparison of conceptual data models. The first category compares one classical data model (for example, a hierarchical or network data model) with another classical data model. A data model is a collection of concepts that can be used to describe a set of data, and operations to manipulate the data. When a data model describes a set of concepts from a given reality, it is called a conceptual data model (Batini et al., 1990). In this paper, a classical model (e.g., hierarchical, network, relational) implies a data model. The second category compares a classical data model with a conceptual data model (for example, EER or SOM). The third compares one conceptual data model with another conceptual data model. Most recent researches are classified within the third category. Table 1 is a summary of past studies. It also compares our current research with that of others.

	Study	Subjects	Data modeling Techniques	[illegible]	Performance
Comparison of Classical Data Models	Brosey and Shneiderman (1978)	Beginner Advanced	Relational Hierarchical	Comprehension Problem solving Memorization	Model Correctness
	Hoffer ([illegible])	Novice	Relational Network Hierarchical	Representation	Database image architecture Confidence Number of files
Comparison of Conceptual Data Model with Classical Data Model	Shoval and Even-Chaime (1987)	IS graduate students	Normalization (Relational) Information analysis (Binary relationship)	Database schema design	Model Correctness Time
	Batra et al. (1990)	Students in introductory MIS course	ER Relational	Modeling	Model correctness Perceived ease-of-use
Comparison of Conceptual Data Models	Bock and Ryan (1993)	Entry-level programmers/Analysts	EER SOM	Modeling	Model correctness
	Kim and March (1995)	Graduate business students IS analysts	EER NIAM	Validation Modeling	Task performance Perceived usefulness
	Hardgrave and Dalal (1995)	Entry-level IS professionals	EER OMT	Understanding data model (Simple, Complex)	Model understanding Understanding time Perceived ease-of-use
	This research	Beginner Expert (Pre-experience on EER) Expert (Inexperience on EER)	EER SOM ORM OMT	Modeling (Natural Language, Form)	Model correctness Understanding time Perceived ease-of-use Effect of pre-EER experience Correlation between given cases

[Table 1] Summary of Literature Review

In past studies, the following is noted. First, there is no evidence that any model fits in all situations. Second, variables are needed that measure the correctness of a model more effectively. Third, we may need a method which can compare a non-object-oriented data model with an object-oriented data model, because object-oriented models rapidly expand their areas and increase their necessities. This paper attempts to answer to these questions.

III. Research Methodology

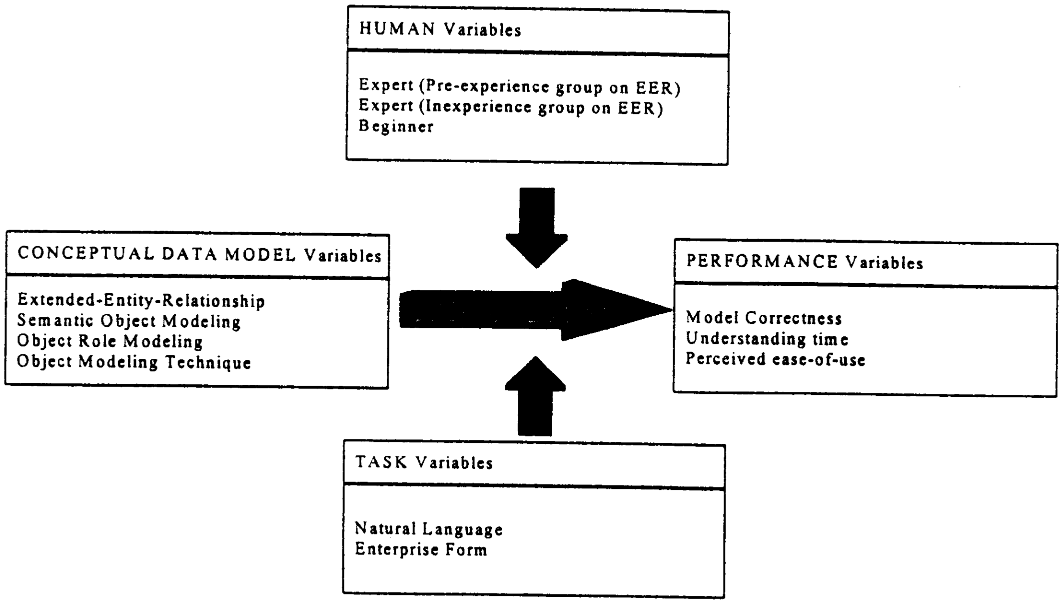
3.1 Research Goal

The type of data modeling technique selected in building a database is very important, because conceptual data modeling techniques serve as a communication vehicle between users and analysts. If we know which modeling techniques are "better," we have many advantages. In many cases, the efficiency of a database is determined at the moment of design. When a database is designed properly, access and maintenance are easy. In spite of the importance of proper database design, there are still no selection criteria for data modeling techniques.

This research provides an empirical study that compares four conceptual data modeling techniques. We distinguish differences of each data modeling techniques in three aspects: model correctness, understanding time, and perceived ease-of-use. As a result of the experiment, we provide some selection criteria for data modeling techniques.

3.2 Research Design

The research model is shown in Figure 1. This research model is modified from the previous studies of data modeling performance (Batra et al., 1990; Bock and Ryan, 1993; Hardgrave and Dalal, 1995). Figure 1 shows the important research question that which data modeling techniques are better than other techniques in each aspect; that is, the focus of this research is on the data model's impact on performance. The data model is independent variable, and performance is a dependent variable, and human and task are control variables.



[Figure 1] Research Model

Four data modeling techniques are regarded as independent variables (EER, SOM, ORM or OMT). EER was selected since it has been widely accepted as a conceptual data modeling technique. SOM was selected because it omitted the concept of relationship in EER. ORM was selected since it used English sentences and omitted the concept of attribute. OMT was selected since it was an enhanced version of EER, in data model view. SOM, ORM and OMT have peculiar features compared with EER, but, have a relationship with EER in some extension.

Human (subject) variable was regarded as a control variable. In order to compare performance among the EER, SOM, ORM, and OMT in the conceptual representation phase of database design, we divided human variable into two groups. One consists of 28 graduates (experts) enrolled in a database system class in KAIST (Korea Advanced Institute Science and Technology). The other consists of 72 undergraduates (beginners) selected from two universities.

The task variable was regarded as a control variable. Task consists of two cases. One is supplied with natural language and the other is with forms. Then natural language case depicts a hospital situation, which is adapted from the Teorey's case (1990). The form case is an actual case of the Miju industry corporation.

Two kinds of case are given to subjects, one natural language and the other forms. We measure performance in each case. Performance in this study is modeling correctness, understanding time, and perceived ease-of-use. Modeling correctness can be defined as the degree to which a conceptual model approaches the correct solution (Batra et al,. 1990). To measure model correctness, all subjects designed each case and measured each facet.

There are many different opinions on overall understanding of conceptual data modeling technique. Batra et al. (1990) and Buck and Ryan (1993) insisted that an overall measure was not necessary or proper, but Hardgrave and Dalal (1995) insisted that it was necessary. In this research, we believe that overall understanding has different meaning to individual facets in the design phase.

Another measure of modeling performance is time (Shoval and Even-Chaime, 1987; Hardgrave and Dalal, 1995). To know the understanding time, we measure total time required in solving the problem. Subjects recorded starting time and finishing time.

The final performance measure is the perceived ease-of use; it was developed by Davis (1989). This measure has been improved and used in many researches (Batra et al., 1990; Hardgrave and Dalal, 1995). This study adopted the measure by Batra et al. (1990).

3.3 Hypothesis

Six different types of hypotheses are tested. The null hypothesis implies that there is no significant difference among the four conceptual data modeling techniques

H1: There is no difference in model correctness among the EER, SOM, ORM, and OMT conceptual data modeling techniques with respect to:

H1a:	Entity/Object/Class
H1b:	Attribute/Entity (Value) type
H1c:	Binary 1:1 relationship
H1d:	Binary 1:M relationship
H1e:	Binary M:M relationship
H1f:	Ternary M:N:O relationship
H1g:	Generalization/Specialization relationship

H2: There is no difference in overall model correctness for relationship among the EER, SOM, ORM, and OMT conceptual data modeling techniques.

H3: There is no difference in understanding time among the EER, SOM, ORM, and OMT conceptual data modeling techniques.

H4: There is no difference in perceived ease-of-use among the EER, SOM, ORM, and OMT conceptual data modeling techniques.

H5: There is no interaction effect between the modeling technique(EER, SOM, ORM, and OMT) and experience for EER(pre-experience, Inexperience)

H6: There is no correlation between tasks (Natural language and Enterprise Form)

3.4 Experiments

The human subjects were divided into four groups: EER, SOM, ORM, and OMT. Subjects were provided with one and a half hours' lectures on each conceptual data modeling technique. The content of the lecture material was designed equivalently for each group. Subject was randomly assigned to one of four treatment groups. They were provided an incentive, educational value. This helped to ensure the high level of motivation that was exhibited by subjects. After one and a half hours' lecture, subjects completed the modeling case.

Error Classification / Item \ Point	Incorrect	Medium Error	Minor Error
	0	0.5	0.75
Entity / Object / Class	•Missing •Represented as an attribute or as a relationship	•Duplicate	•Extra entity
Attribute / Entity type, Value type	•Missing	•Incorrect connectivity •Use of an existing entity / object / class name	•Duplicate •Incorrect symbol
Relationship	•Missing •Incorrect degree	•Incorrect connectivity	•Incorrect name
Generalization	•Missing	•Incorrect inheritance	•Incorrect symbol

[Table 2] Grading Scheme

We modified the grading scheme used by Batra et al. (1990), Bock and Ryan (1993), and Kim and March (1995). The score varies from zero points for "completely incorrect" to one point for "completely correct". Scores with medium or minor errors are 0.50 and 0.75 points, respectively. Table 2 shows this grading scheme.

IV. Result Interpretation

Human subjects consist of 28 graduates and 72 undergraduates. In the undergraduate group, one subject (in SOM), and two subjects (in ORM and OMT), did not respond. In the graduate group, all subjects responded.

4.1 Model Correctness, Understanding Time, and Perceived Ease-of-Use

When a modeling task is given to beginners in natural language, the results of model correctness, understanding time and perceived ease-of-use are shown as Table 3. This shows that there are significant differences among the data modeling techniques in entity/object/class, binary 1:1, binary M:M, ternary M:N:O, generalization, total relationship, understanding time, and perceived ease-of-use. In particular, binary 1:1 relationship score is zero except EER. In Table 3, the shaded boxes are one group and the normal (unshaded) boxes are another group. For example, SOM and ORM is the same group, and EER and OMT is another in entity/object/class. From the result of the Duncan test (Hair et al., 1995), EER, SOM, ORM, and OMT are divided into two group in all significantly different cases.

Model / Facet	EER		SOM		ORM		OMT		P-value
	Mean	Stdev	Mean	Stdev	Mean	Stdev	Mean	Stdev	
Entity/Object/Class	9.79	1.71	4.27	0.88	[illegible]	3.45	8.85	1.98	0.0000 *
Attribute/Entity(Value) type	12.43	1.37	10.79	2.99	11.20	2.99	11.20	12.03	0.1669
Binary 1:1 relationship	0.277	0.46	0.00	0.00	0.00	0.00	[illegible]	0.00	0.0017 *
Binary 1:M relationship	1.11	0.83	0.64	0.70	0.73	0.96	1.13	0.83	0.2239
Binary M:M relationship	0.44	0.51	0.00	0.24	[illegible]	0.00	0.33	0.48	0.0026 *
Ternary M:N:O relationship	0.77	0.42	0.00	0.00	[illegible]	0.51	0.46	0.53	0.0000 *
Generalization relationship	0.81	0.38	0.00	0.00	[illegible]	0.36	0.60	0.50	0.0000 *
Total Relationship Score	3.43	1.46	0.70	0.84	[illegible]	1.45	2.53	1.80	0.0000 *
Understanding Time	127.7	24.1	140.0	23.7	[illegible]	30.6	126.0	31.8	0.0103 *
Perceived Ease-of-Use	19.44	6.08	15.05	6.23	13.80	5.24	17.33	4.8	0.0286 *
Number of Respondents	(N:18)		(N:17)		(N:15)		(N:15)		N/A

Note: ✱: Significant difference at level of $\alpha = 0.05$.

[Table 3] Comparison Summary of Beginner in Natural Language

When we give a modeling task which is written in natural language to experts, the results of model correctness, understanding time and perceived ease-of-use are like Table 4, which shows that there are significant differences among the data modeling techniques in binary 1:1, generalization, total relationship, understanding time, and perceived ease-of-use. Specially, Table 4 shows that SOM scores are very low, compared with those in other data modeling techniques.

Model / Facet	EER		SOM		ORM		OMT		P-value
	Mean	Stdev	Mean	Stdev	Mean	Stdev	Mean	Stdev	
Entity/Object/Class	10.92	0.12	9.78	1.72	10.78	0.39	10.57	1.02	0.1867
Attribute/Entity(Value) type	13.25	1.08	13.92	1.05	13.28	0.87	12.50	3.57	0.6201
Binary 1:1 relationship	0.85	0.37	0.00	0.00	0.42	0.53	0.71	0.48	0.0034 *
Binary 1:M relationship	1.71	0.48	1.57	0.53	1.57	0.53	1.85	0.37	0.6530
Binary M:M relationship	0.35	0.47	0.57	0.53	0.85	0.37	0.57	0.53	0.3127
Ternary M:N:O relationship	0.28	0.48	0.00	0.00	0.14	0.37	0.42	0.53	0.2590
Generalization relationship	0.85	0.37	0.14	0.37	1.00	0.00	1.00	0.00	0.0000 *
Total Relationship Score	4.07	1.23	2.28	2.28	4.00	1.00	4.57	1.13	0.0051 *
Understanding Time	157.8	63.8	251.4	143.1	218.5	80.5	124.2	56.0	0.0675 **
Perceived Ease-of-Use	21.14	4.63	23.28	8.40	11.85	2.47	20.57	2.37	0.0017 *
Number of Respondents	(N:7)		(N:7)		(N:7)		(N:7)		N/A

Note : ✱✱: Significant difference at level of $\alpha = 0.1$.

[Table 4] Comparison Summary of Expert in Natural Language

When we give a modeling task which is made of enterprise form to beginners, the results of model correctness, understanding time and perceived ease-of-use are shown as Table 5.

Model / Facet	EER		SOM		ORM		OMT		F-value
	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.	
Entity/Object/Class	6.13	2.09	4.64	1.10	5.40	1.09	5.43	5.43	0.0331 *
Attribute/Entity (Value) type	19.06	5.09	16.75	4.33	18.71	3.76	18.45	4.45	0.4440
Binary 1:1 Relationship	0.11	0.32	0.00	0.00	0.00	0.00	0.00	0.00	0.1497
Binary 1:M Relationship	1.88	0.75	0.94	0.89	1.46	0.91	1.86	0.74	0.0045 *
Binary M:M Relationship	0.61	0.50	0.23	0.43	[illegible]	0.45	0.53	0.51	0.0612 **
Ternary M:N:O Relationship	0.11	0.32	0.05	0.24	0.00	0.00	0.00	0.00	0.3701
Generalization Relationship	0.20	0.40	0.05	0.24	0.06	0.25	0.13	0.35	0.5022
Total Relationship Score	2.93	1.11	1.29	1.40	1.80	1.26	2.53	1.18	0.0013 *
Understanding Time	133.3	29.3	157.7	51.3	121.6	52.2	132.0	28.8	0.0395 *
Perceived Ease-of-Use	18.38	4.16	13.52	5.06	13.53	5.48	17.64	4.68	0.0057 *
Number of Respondents	(N:18)		(N:17)		(N:15)		(N:15)		N/A

[Table 5] Comparison Summary of Beginner in Form

Table 5 shows that there are significant differences in entity/object/class, binary 1:M, binary M:M, total relationship, understanding time, and perceived ease-of-use among the data modeling techniques. In particular, the binary 1:1 relationship score is zero except EER, which is the same result of natural language.

When we give a modeling task which is made of enterprise form to experts, the results of model correctness, understanding time and perceived ease-of-use are like those shown in Table 6.

Model / Facet	EER		SOM		ORM		OMT		F-value
	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.	
Entity/Object/Class	8.60	0.31	8.00	0.94	8.60	0.28	8.25	0.52	0.1694
Attribute/Entity (Value) type	28.07	0.89	28.32	1.29	23.53	1.90	26.0	3.46	0.0009 *
Binary 1:1 Relationship	0.71	0.48	0.00	0.00	0.57	0.53	0.85	0.37	0.0036 *
Binary 1:M Relationship	2.42	1.13	3.00	0.81	3.14	0.89	2.64	0.47	0.4098
Binary M:M Relationship	0.42	0.53	0.85	0.89	0.14	0.37	1.00	0.81	0.1056
Ternary M:N:O Relationship	0.14	0.37	0.57	0.53	0.00	0.00	0.00	0.00	0.0095 *
Generalization Relationship	0.57	0.53	0.28	0.48	0.92	0.18	0.85	0.37	0.0327 *
Total Relationship Score	4.28	1.97	4.71	1.97	4.78	0.80	5.35	1.24	0.6585
Understanding Time	120.7	53.8	229.2	130.3	164.2	34.0	125.0	63.5	0.0604 **
Perceived Ease-of-Use	23.57	4.39	24.28	8.86	15.00	3.74	22.0	2.23	0.0142 *
Number of Respondents	(N:7)		(N:7)		(N:7)		(N:7)		N/A

[Table 6] Comparison Summary of Expert in Form

Table 6 shows that there are significant differences in attribute/entity (value) type, binary 1:1, ternary M:N:O, generalization, total relationship, understanding time, and perceived ease-of-use among the data modeling techniques. Especially, the binary 1:1 relationship score is zero in SOM.

4.2 Interaction Effect between Data Model and Experience on EER

In the natural language case, the results of interaction effect between data model and experience on EER are shown as Table 7.

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	10.0	4.99	6.92	9.10	11.0	10.3	10.7	10.41	0.056 *
[illegible]	0.42	0.00	0.00	0.06	1.00	0.0	0.50	0.80	0.027 *
[illegible]	135	167	153	130	150	197	236	108	0.096 **
[illegible]	19.8	15.4	13.7	17.7	22.0	27.5	11.6	20.6	0.005 *

[Table 7] Interaction Effect between Data Model and Experience on EER in Natural Language

Table 7 shows that there are significant differences in entity/object/class, binary 1:1 relationship, time, and perceived ease-of-use between pre-experience in the EER group and inexperience in the EER group. Pre-experience in the EER group is better than inexperience in the EER group, in most cases, but in case of ORM, inexperience in the EER group obtains a higher score in perceived ease-of-use than pre-experience in the EER group. As result of this feature, ORM may be easier for beginners.

In the enterprise form case, the results of interaction effect between data model and experience on EER are like those in Table 8.

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	0.25	0.00	0.06	00.12	1.00	0.0	0.50	0.80	0.042 *
[illegible]	2.00	1.15	1.50	2.00	3.00	3.50	3.33	2.50	0.031 *
[illegible]	0.08	0.10	0.00	0.00	1.00	0.75	0.00	0.00	0.000 *
[illegible]	19.8	26.0	13.1	18.0	20.0	28.0	15.6	22.0	0.017*

[Table 8] Interaction Effect between Data Model and Experience on EER in Form

In Table 8, there are significant differences in binary 1:1 relationship, binary 1:M relationship, ternary M:N:O, and perceived ease-of-use between pre-experience in the EER group and inexperience in the EER group. In contrast to natural language in all facets, pre-experience in the EER group is better than inexperience in the EER group in form case.

4.3 Correlation between Natural Language Case and Enterprise Form Case

In case of beginners and experts , the results of correlation between natural language case and enterprise form case are shown as Table 9 and Table 10.

[illegible]	[illegible]
Entity/Object/Class	0.2443 (0.050)*
Attribute/Entity(Value) type	0.2192 (0.079)**
Binary 1:1 relationship	0.6172 (0.000) *
Time	0.8388 (0.000)*
Perceived ease of use	0.7917 (0.000)*

[Table 9] Summary of Correlation between Natural Language and Form in Beginner

[illegible]	[illegible]
Entity/Object/Class	0.6668 (0.000)*
Attribute/Entity(Value) type	0.4698 (0.012)*
Binary 1:1 relationship	0.5013 (0.007) *
Time	0.7617 (0.000)*
Perceived ease of use	0.8485 (0.000)*

[Table 10] Summary of Correlation between Natural Language and Form in Expert

Table 9 and Table 10 show that there are correlations between natural language case and enterprise form case in entity/object/class, attribute/entity (value) type, binary 1:1, time, and perceived ease-of-use, without regard to subjects. As result of this experiment, the better subjects in the natural language case may be also better subjects in form case in the entity/object/class, attribute/entity (value) type, binary 1:1, time, and perceived ease-of-use, but we cannot sure that the better subjects in the natural language case will be also better in the form case.

4.4 Result Summary

In summary, a comparison among EER, SOM, ORM, and OMT shows some interesting results as follows.

Task / Criteria	Natural Language		Form	
	Relatively Better	Relatively Worse	Relatively Better	Relatively Worse
Entity/Object/Class	EER, OMT	SOM,ORM	EER	SOM,ORM,OMT
Attribute/Entity (Value) type	EER, SOM, ORM, OMT†		EER, SOM, ORM, OMT†	
Binary 1:1	EER	SOM,ORM, OMT		EER, SOM, ORM, OMT†
Binary 1:N	EER, SOM, ORM, OMT†		EER, OMT	SOM,ORM
Binary M:N	EER, OMT	SOM,ORM	EER, OMT	SOM,ORM
[illegible]	EER, OMT	SOM,ORM		EER, SOM, ORM, OMT†
Generalization	EER, OMT	SOM,ORM		EER, SOM, ORM, OMT†
Total Relationship Score	EER, OMT	SOM,ORM	EER, OMT	SOM,ORM
Understanding Time	EER, SOM, OMT	ORM	EER, OMT	SOM,ORM
Perceived Ease-of-use	EER, OMT	SOM,ORM	EER, OMT	SOM,ORM

[Table 11] Result Summary of Beginners

Task / Criteria	Natural Language		Form	
	Relatively Better	Relatively Worse	Relatively Better	Relatively Worse
Entity/Object/Class	EER, SOM, ORM, OMT†		EER, SOM, ORM, OMT†	
Attribute/Entity (Value) type	EER, SOM, ORM, OMT†		EER, SOM,OMT	ORM
Binary 1:1	EER, OMT	SOM,ORM	EER,ORM, OMT	SOM
Binary 1:N	EER, SOM, ORM, OMT†		EER, SOM, ORM, OMT†	
Binary M:N	EER, SOM, ORM, OMT†		EER, SOM, ORM, OMT†	
[illegible]		EER, SOM, ORM, OMT†	EER, SOM	ORM, OMT
Generalization	EER, ORM, OMT	SOM	EER, ORM, OMT	SOM
Total Relationship Score	EER, ORM, OMT	SOM	EER, SOM, ORM, OMT†	
Understanding Time	EER, ORM, OMT	SOM	EER, ORM, OMT	SOM
Perceived Ease-of-use	EER, SOM, OMT	ORM	EER, SOM, OMT	ORM

[Table 12] Result Summary of Experts

First, binary 1:1 relationship is difficult for beginners to design without regard to data modeling techniques. In experts, SOM is difficult in presenting binary 1:1 relationship. Low score of binary 1:1 relationship in SOM may be due to less concrete modeling methods. The symbol of binary 1:1

relationship of EER, ORM, and OMT adds another graphical notation to the same entity/object, but SOM requires the addition of another SOL (Semantic Object Link) to the same object.

Second, the scores of ternary relationship of SOM and ORM are very low in beginners, without regard to cases. In experts, SOM has a high score compared with those of other modeling techniques in form case.

Third, in the natural language case, the score of generalization relationship of beginners is very low in SOM and ORM. In the form case, the score of generalization relationship of beginners is very low, and the experts have difficulty in presenting a generalization relationship of SOM in both cases.

Fourth, ORM requires much time for beginners. Data modelers would feel that ORM is more difficult to use, and may make more efforts to learn ORM.

Fifth, there is no significant difference between pre-experience in the EER group and inexperience in the EER group except binary 1:1 relationship. ORM may also be more appropriate for beginners than experts. Inexperienced modelers on EER may feel that ORM is easier because it employs English sentences.

Finally, designers who model well in the natural language case do so in enterprise form case in entity/object/class, attribute/entity (value) type, binary 1:1, time, and perceived ease-of-use. In case of time and perceived ease-of-use, a high correlation between natural language and enterprise form is noted.

V. Conclusions

This study has empirically compared four conceptual data modeling techniques such as EER, SOM, ORM and OMT with respect to three dimensions: (i) model correctness, (ii) time to understand, and (iii) perceived ease-of-use. Cases in natural language or in enterprise form are given for data modeling. Human subjects are categorized into beginners and experts.

Results indicate some significant differences among four techniques. For beginners, EER and OMT would produce more correct designs than SOM and ORM. For experts, SOM would produce less correct designs especially for the case in natural language. Similarly, EER and OMT are faster in time to understand. ORM would be more difficult to use than other modeling techniques.

Our empirical study does not answer why one model is better than another. However, the above results are positive findings because they may help data modelers better understand different modeling techniques.

References

[1] Batini, C., Ceri, S., & Navathe, B.S. (1992). Conceptual Database Design: An Entity-Relationship approach. The Benjamin/Cummings Publishing Company Inc.

[2] Batra, D, Hoffer, J.A., & Bostrom, R.P. (1990). Comparing representations with relational and EER models. Communications of the ACM. 33(2), 126-139.

[3] Blaha, M.R., Premerlani, W.J., & Rumbauch, J.E. (1988). Relational database design using an object-oriented methodology. Communications of the ACM. 31(4), 414-427.

[4] Bock, D.B., & Ryan, T. (1993). Accuracy in Modeling with Extended Entity-Relationship and Object Oriented Data Model. Journal of Database Management. 4(4), 30-39.

[5] Brosey, M., & Shneiderman, B. (1978) Two experimental comparisons of relational and hierarchical database model. International Journal of Man-Machine Studies 10, 625-637.

[6] Chen, P.P. (1976). The entity-relationship model: Toward a unified view of data. ACM Transactions on Database Systems. 1(1), 9-36.

[7] Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly. 13(3), 319-340.

[8] Eckert, G., & Golder, P. (1994). Improving object-oriented analysis, Information and Software Technology. 36(2), 67-86

[9] Hair, J.F., Anderson, R.E., Tatham, R.L., & Black, W.C. (1995). Multivariate Data Analysis, 5th Edition. Prentice Hall Inc.

[10] Halpin, T. (1995). Conceptual Schema and Relational Database Design, 2nd Edition. Prentice Hall Inc.

[11] Hardgrave, B.C., & Dalal, N. P. (1995). Comparing object-oriented and extended entity-relationship data models. Journal of Database Management. 6(3), 15-21.

[12] Hoffer, J.A. (1982). An empirical investigation into individual differences in database models. In Proceedings of the Third International Conference on Information Systems. 153-168.

[13] Kim, Young-Gul & March, S.T. (1995). Comparing data modeling formalism. Communication of the ACM. *38(6)*, 103-114.

[14] Kroenke, D.M. (1995). Database Processing Conceptual Database Design, 5th Edition. Prentice Hall Inc.

[15] Nijssen, G.M. (1994). A general analysis procedure: Recent advances in universal informatics. Proceeding 2nd NIAM-ISDM conference. Albuquerque, NM.

[16] Rumbauch, J., Blaha, M., Premerlani, W., Eddy, F., & Loresen., W. (1991) Object-Oriented Modeling and Design. Prentice Hall, Englewood Cliffs, N. J.

[17] Shoval, P., & Even-Chaime, M. (1987). Database schema design: An experimental comparison between normalization and information analysis. Database. 18(3), 30-39.

[18] Teorey, T.J. (1990). Database Modeling and Design: The Fundamental Principles, 2nd Edition. Morgan Kaufmann Publishers Inc.

[19] Teorey, T.J., Yang, D., & Fry, J.P. (1986). A logical design methodology for relational databases using the extended entity-relationship model. ACM Computing Survey. 18(2), 197-222.