

CONCEPTUAL DATA MODELLING: AN EMPIRICAL STUDY OF EXPERT AND NOVICE DATA MODELLERS

Graeme Shanks
Department of Information Systems
Monash University
Melbourne, Australia
Email: graeme.shanks@is.monash.edu.au

ABSTRACT

This study explores the differences between conceptual data models designed by expert and novice data modelling practitioners. The data models are evaluated using a number of quality factors synthesised from previous empirical studies and frameworks for quality in conceptual modelling. This study extends previous studies by using practitioners as participants and using a number of different quality factors in the evaluation. The study found that data models produced by expert data modellers are more correct, complete, innovative and flexible than those produced by novices. The results suggest that further research into the aspects of expertise that lead to such differences and how training courses can narrow the gap between expert and novice performance is required.

INTRODUCTION

Conceptual data models provide a "precise and unambiguous representation of organisational information requirements" (Kim and March 1995, p103). They also provide a means of communication amongst information systems professionals and business users when discovering and modelling information requirements. Entity-Relationship (ER) models are a widely used form of conceptual data model (Chen 1976).

The development of a conceptual data model is a design activity influenced by various kinds of knowledge including both application domain knowledge and data modelling representation and process knowledge. When designing a conceptual data model, expert data modellers develop a holistic understanding of a problem, categorise problem descriptions into standard abstractions and reuse generic data models from their previous experience (Batra and Davis 1992, Chaiyasut and Shanks 1994). Novice data modellers follow simple guidelines such as "make a list of all the nouns ..." (Veryard 1984, p15). Many previous empirical studies of conceptual data modelling have focused on comparing representation formalisms or have traced the process of the modelling task.

This study explores the differences between conceptual data models designed by expert and novice data modelling practitioners using the entity relationship modelling formalism. The data models are evaluated using a number of quality factors synthesised from previous empirical studies and frameworks for quality in conceptual modelling. This study extends previous studies by using practitioners as participants and using a number of different quality factors in the evaluation. The results suggest that further research into the aspects of expertise that lead to such differences and how training courses can narrow the gap between expert and novice performance is required.

The paper is structured as follows. The next section of the paper reviews previous empirical studies in conceptual data modelling. The third section describes the research approach adopted in the study. The following section discusses the results of the study and the final section of the paper presents a number of implications of the study and suggests areas for further research.

PREVIOUS EMPIRICAL STUDIES IN CONCEPTUAL DATA MODELLING

There have been a number of previous empirical studies of conceptual data modelling. These studies have investigated the effects of different data modelling formalisms (Juhn and Naumann 1985, Jarvenpaa and Machesky 1986, Shoval and Even-Chaime 1987, Batra *et al.* 1990, Kim and March 1995) and expert and novice performance in conceptual data modelling in terms of both product and process (Batra and Davis 1992, Maiden and Sutcliffe 1992, Shanks *et al.* 1993, Chaiyasut and Shanks 1994). These studies vary in the participants involved, the data modelling formalisms used, the kinds of task undertaken and the focus of the study. A comparative summary of these studies is presented in Table 1 (adapted and extended from Kim and March 1995, p104).

It can be seen that many of the participants involved in the studies were students, the most commonly used data modelling formalism was entity relationship modelling, the most common task was to build a conceptual data model from a small case description and the focus of most studies was on the quality of the conceptual data model designed.

Table 1 Summary of Previous Empirical Studies in Conceptual Data Modelling

<i>Study</i>	<i>Participants</i>	<i>Data Modelling Formalisms</i>	<i>Task</i>	<i>Focus</i>
Juhn and Naumann (1985)	MIS MBA students (novices)	ER (and LDS) Relational (and DAD)	Comprehension, Modelling	Quality of data model
Jarvenpaa and Machesky (1986)	IS Undergrad. students (analysts)	LDS Relational	Modelling	Quality of data model
Shoval and Even-Chaime (1987)	IS Grad. students (analysts)	Object-role Relational	Modelling	Quality of data model, Modelling time
Mantha (1987)	Practitioners (analysts)	ER	Modelling	Quality of data model
Batra <i>et al.</i> (1990)	IS Undergrad. students (novices)	ER, Relational	Modelling	Quality of data model
Batra and Davis (1992)	Students and Practitioners, (Expert/Novice)	ER	Modelling	Cognitive Process
Maiden and Sutcliffe (1992)	Grad. IS students (novices)	DFD	Modelling	Cognitive Process, Quality of data model
Shanks <i>et al.</i> (1993)	Grad. IS students and IS practitioners (Expert/novice)	ER	Modelling	Quality of data model
Chaiyasut and Shanks (1994)	IS Practitioners (Expert/novice)	ER	Modelling	Cognitive Process
Kim and March (1995)	Grad. IS students and IS practitioners (analysts)	EER, Object-role	Modelling, Model Validation	Comprehension, Quality of data model

Those studies which investigated the effects of different data modelling formalisms found that semantic data models (such as ER) lead to the design of better quality data models than the formalisms based on storage structures (such as relational). Other studies traced the cognitive process of expert and novice data modellers in the conceptual data modelling task. These studies showed that expert data modellers were better able to conceptualise and understand the case description than novices. This enabled them to partition the case description into sub-problems, reuse patterns from their previous experience, spend more time evaluating their solutions and move their solutions to higher levels of abstraction.

Both Maiden and Sutcliffe (1992) and Shanks *et al.* (1993) evaluated the quality of conceptual data models designed by expert and novice data models. In each case the models developed by experts were found to be of higher quality than those of novices. This study extends these previous studies in several ways:

- All participants in the study are information systems practitioners rather than students. This should increase the external validity of the results (Hoffer 1982).
- The evaluation of the models was more comprehensive and based on several quality factors synthesised from a number of quality frameworks.
- The ER formalism used was the Martin (1987) standard, which is widely used in practice, rather than the Chen standard commonly used in other studies (Hitchman (1995) found that only 5% of practitioners used the Chen notation).

RESEARCH APPROACH

The research approach used is a "differential study" (Graziano and Raulin 1989). A differential study is in many ways similar to an experiment although in a true experiment participants are randomly assigned to groups. In a differential study groups are formed from values of pre-existing variables.

Research Model

The research model for this study is shown in Figure 1 and depicts the relationships between the participants, the task and the quality of the data models.

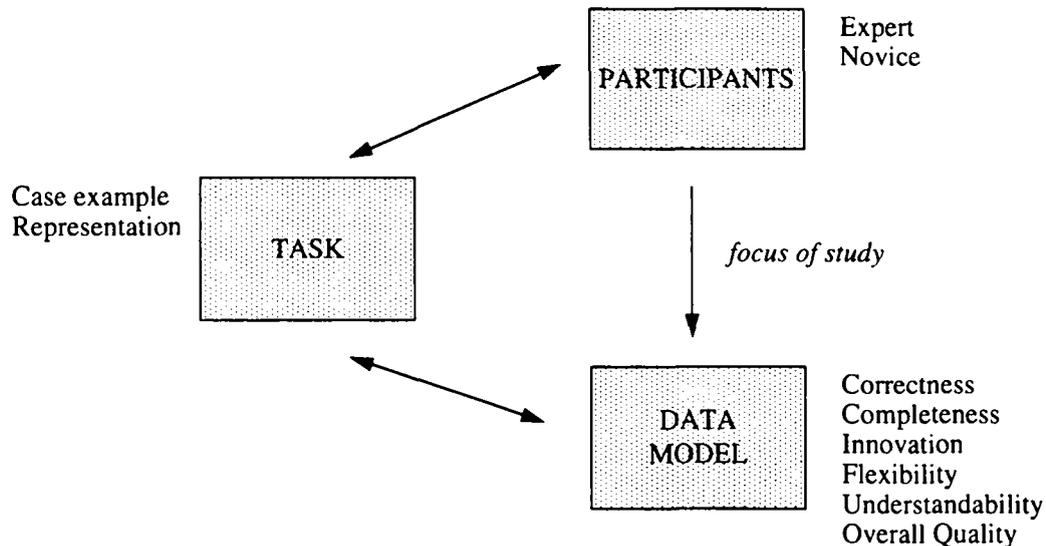


Figure 1 Research Model

The Task

Each participant in the study was required to read a case study description and prepare a conceptual data model. The case study problem was based on a small electronics trading and repair company. The description of the case study problem was provided by the manager of the company using a tape recorder and later transcribed into written form. Participants worked entirely from the narrative; there was no "user" present to answer questions. Because the case example described a common, familiar type of company there was no need to have any specialised domain knowledge to understand its operations. Each participant was also required to complete a brief questionnaire after the data modelling task containing details of their industry experience and training. Completion of the task and questionnaire took between one and two hours.

The Participants - Independent Variable

The participants in this study were all practising information systems professionals. Each of the participants had attended a three day course in entity relationship modelling. Expertise in data modelling is defined by the number of years of experience as a specialist data modeller together with the number of data models to which the participant contributed. Participants with at least four years experience as specialist data modellers and who had contributed to the development of at least ten conceptual data models were categorised as expert. Other participants were categorised as novice. It can be seen that the definition of an expert data modeller is very stringent.

Data Model Quality - Dependent Variable

Quality is defined as "...an emergent property measuring the extent to which the [model] is fit for use and meets the customer's expectations" (Kaposi and Kitchenham 1987, p3). Moody and Shanks (1994) note that there are two aspects to understanding quality in conceptual data modelling; process quality

and product quality. Process quality relates to understanding and improving the process of building a conceptual data model. Product quality refers to understanding and evaluating the data model which is built. This study focuses on product quality.

A number of frameworks have been proposed which can be used to understand and evaluate quality in conceptual data models (see for example Shanks *et al.* 1993, Lindland *et al.* 1994, Krogstie *et al.* 1995, Moody and Shanks 1994 and Simsion 1994). In addition a number of previous studies have developed criteria for comparing conceptual data models using different notations (see for example Batra *et al.* 1990, Mantra 1987, Shoal and Even-Chaime 1987, Kim and March 1995, Hoffer 1982). The quality factors used in this study are synthesised from these previous studies and include correctness, completeness, innovation, flexibility and understandability. In addition, an "overall quality" rating is obtained for each model. This is to ensure that "transcendental" qualities which cannot be factored out are considered (Kaposi and Kitchenham 1987). Each of these quality factors and its evaluation is defined below.

A combination of objective measures and subjective ratings is necessary when measuring quality within the software domain (Kaposi and Kitchener (1987). Subjective quality ratings can be strengthened in three ways (Moody and Shanks 1994):

- By providing clear and precise definitions of the quality criteria
- By using many reviewers and applying inter-rater reliability measures
- By getting expert and experienced reviewers to carry out the assessment.

A definition for each quality criteria used in this study and a discussion of its measurement follows.

(a) Correctness

The correctness of a conceptual data model concerns its syntactic representation (Lindland *et al.* 1994). Correctness is defined as the degree to which the data model conforms to the syntax rules of the particular notation used. All participants were trained in the variant of the entity relationship notation defined in Martin (1987). The evaluation instrument used for correctness is based on those developed by Ridjanovic (1986), Batra *et al.* (1990) and Kim and March (1995). Evaluation of correctness is based on the number of major and minor errors found in the data models. Major errors include no name for entities or relationships, no cardinality constraint on relationships and no optional/mandatory symbol on relationships. Minor errors include duplicate entity or relationship names, non-noun entity names and incorrect symbols. A major syntactic error is assigned a 0.5 penalty and a minor syntactic error is assigned a 0.25 penalty (Kim and March 1995). The algorithm for correctness (expressed as a percentage) is:

$$\text{Correctness} = \frac{N - 0.5E_1 - 0.25E_2}{N} \times 100$$

where N is the number of instances of constructs (entities and relationships) in the model, E₁ is the number of major errors for each construct and E₂ is the number of minor errors for each construct. The overall correctness is obtained by averaging the participant's performances for each modelling construct.

Because all participants have had equivalent training and the entity relationship notation has a relatively simple syntax, it is expected that there should not be any significant differences in data model correctness between the two groups. Hence the following hypothesis is posited:

- H1 There will be no difference in the correctness of data models built by expert data modellers from those built by novice data modellers.

(b) Completeness

The completeness of a conceptual data model concerns its perceived semantic content (Krogstie *et al.* 1995). Completeness is defined as the extent to which the data model supports user information requirements. The evaluation instrument used for completeness is based on those developed by Ridjanovic (1986), Batra *et al.* (1990) and Kim and March (1995). Evaluation of completeness is based on a comparison of each of the data models with a "correct" solution developed by the author and another data modelling academic. The "correct" solution determines the features which should be present in each of the data models. Kim and March (1995 p109) note that "... a required construct is considered to be present if there is any semantically equivalent construct in the participant's data model. For example, a concept represented by an entity in the "correct" model could be represented as ... a relationship in the [participants] model"

Evaluation of completeness is based on the number of major and minor errors found in the data models. Major errors include missing entities or relationships. Minor errors include incorrect extra or redundant entities or relationships, incorrect cardinality and incorrect degree. A major syntactic error is assigned a 1.0 penalty and a minor syntactic error is assigned a 0.3 penalty (Kim and March 1995). The algorithm for completeness (expressed as a percentage) is:

$$\text{Completeness} = \frac{N - E_1 - 0.3E_2}{N} \times 100$$

where N is the number of constructs (entities and relationships) in the "correct" model, E_1 is the number of major errors for each construct and E_2 is the number of minor errors for each construct. The overall completeness is obtained by averaging the participant's performances for each modelling construct.

It is expected that the data models designed by experts should be more complete than the data models developed by novices. Hence the following hypothesis is posited:

H2 Expert data modellers will design data models of higher completeness than novice data modellers.

(c) Innovation

The level of innovation in a conceptual data model is defined as the extent to which "new concepts" are introduced into the model (Shanks *et al.* 1993). These "new concepts" must be relevant and valid for the model. A data model is said to be "literal" if it only includes concepts directly mentioned by the user. In this study a data model which included only entity types which matched nouns in the case description was deemed to exhibit no innovation. Those data models which included entity types which were not explicitly included in the case description were considered innovative.

In order to evaluate innovation, a list of nouns was developed for the case example narrative. A count was made of the number of entity types in each of the data models which did not correspond to nouns in the noun list. Those entity types which were clearly only resolving many-to-many relationships and did not add any new concepts to the models were not counted. The ratio of "innovative" entity types to the total number of entity types in each data model was used as a measure of the level of innovation in the data models. The algorithm for innovation (expressed as a percentage) is:

$$\text{Innovation} = \frac{N_i}{N_t} \times 100$$

where N_i is the number of "innovative" entity types and N_t is the total number of entity types.

Previous empirical studies have shown that experienced data modellers recognise similar problem categories and are able to reuse data model patterns (Batra and Davis 1992, Chaiyasut and Shanks 1994). These patterns may include entity types not included in the case example narrative and thus will be seen as innovations. Experienced data modellers frequently use the aggregation and generalisation abstractions to introduce new concepts to data models (Chaiyasut and Shanks 1994, Simsion 1994). Hence the following hypothesis is posited:

H3 Expert data modellers will design data models with higher innovation than novice data modellers.

(d) Flexibility

Flexibility is defined as the ease with which the data model can reflect changes in requirements without changing the data model itself. Other terms for flexibility include extensibility, stability and evolvability (Moody and Shanks 1994). Flexibility is one of the most important characteristics of a data model (Avison and Fitzgerald 1995, Leviten and Redman 1995) and has a major effect on the maintenance costs of a system (Moody and Shanks 1994). The use of generic concepts and the representation of business rules as structures in the data model will lead to a more flexible data model (Barker 1989, Simsion 1994).

There is no objective measure for the flexibility of a data model. Flexibility is given subjective expert ratings using a seven point likert scale. The following hypothesis is posited:

H4 Expert data modellers will design data models with higher flexibility than novice data modellers.

(e) Understandability

Understandability relates to the ease with which concepts and structures in the data model can be understood by stakeholders in the data modelling process. Each stakeholder will interpret and understand the model in different ways (Krogstie *et al.* 1995). The ratings reflect the ability of the reviewers to understand the data models and were given using a seven point likert scale.

Because expert data modellers are expected to produce more correct and complete data models, it is expected that the data models designed by experts should be more understandable than the data models developed by novices. Hence the following hypothesis is posited:

H5 Expert data modellers will design data models that are more understandable to the reviewers than the models of novice data modellers.

(f) Overall Quality

Overall quality is a measure of the raters' subjective evaluation of the model in total. This measure is included to account for the "transcendental" properties of the data models which are not included in the other quality factors (Kaposi and Kitchenham 1987). Alexander (1979) refers to these properties as "the quality without a name". It is expected that the data models designed by experts should have an overall quality greater than the data models developed by novices. Hence the following hypothesis is posited:

H6 Expert data modellers will design data models of higher overall quality than novice data modellers.

Data Analysis and Selection of Statistical Tests

Each of the three reviewers of the models was an academic specialising in the area of conceptual data modelling who also had previous industry experience in conceptual data modelling. However, none of the reviewers would have met the criteria for inclusion in the expert data modeller category. To enhance the validity of the evaluation all data models were transcribed using the same template so that the evaluation would focus on only the content of the data models rather than the presentation. To avoid sequencing bias each of the raters evaluated the models in a different sequence. Inter-rater reliability (Cronbach Alpha co-efficient) was determined for all subjective ratings and for any scoring which involved expert judgement.

The data collected from the evaluation is mostly interval data (correctness, completeness and innovation). Likert scale ratings for understandability and flexibility are ordinal, however the treatment of ordinal data as interval data is not uncommon in information systems and social science research (Periasamy 1994, Pervan and Klass 1992). The seven point likert scales had literal descriptions of only the mid-point and end-points thus providing no descriptive influence on the distances between the points on the scale. Accordingly, mean, standard deviation and t-test were used for statistical inference and correlation was used for exploratory data analysis. The Statview statistical software tool was used for all statistical analysis.

RESULTS**Participant Characteristics**

A summary of participant characteristics is shown in Table 2. It can be seen that the categorisation of participants as expert or novice (using the criteria defined in section 3.3) agrees well with their self assessment.

Table 2 Summary of Participant Characteristics

	EXPERT: 18 participants	NOVICE: 21 participants
Years Data Modelling Experience (mean/stdev)	6.41 (2.85)	1.52 (0.93)
Number of Data Models	Greater than ten 18 Three to ten 0 One or Two 0 None 0	Greater than ten 0 Three to ten 3 One or Two 12 None 6
Practical Expertise (Self Assessment)	Expert 13 Novice 5	Expert 0 Novice 21
Theoretical Expertise (Self Assessment)	Expert 11 Novice 7	Expert 0 Novice 21

Hypothesis Testing

Results for each of the six hypotheses tested are summarised in Table 3. For all hypotheses, tests for significant differences between the two groups were conducted. In addition, for understandability and flexibility the mean value was tested for significant difference from a normally distributed population with a mean value of 4.0 (mid-point on likert scale).

The first hypothesis was rejected ($p=0.0137$, significant with $\alpha = 0.05$) with high inter-rater reliability. Expert data modellers produced data models which were more correct than data models produced by novice data modellers. This result could be explained by the more frequent use of the data modelling notation by the expert data modellers despite both groups receiving the same recent three day training course. Considering the simple syntax of the entity relationship model, the average scores for both experts and novices were quite low. In order to explore whether any particular facet of the notation contributed most to the difference a second scoring of the data models was conducted (see Table 4). Only 53% of expert data modellers and 29% of novice data modellers used relationship names. This is of great concern as relationship names are essential in documenting the business rule which relationships represent. A possible explanation is that the notation used in this study uses lines to represent relationships. Representing relationships as diamonds may increase their prominence in the model and encourage naming.

Table 3 Summary of Hypothesis Testing Results

Hypothesis	Expert (mean/stdev)	Novice (mean/stdev)	Inter-rater Reliability (alpha)	T-value	P-value
Correctness (%)	84.33 (17.24)	71.52 (17.47)	0.97	$t(37)=2.29$	0.0137
Completeness (%)	72.89 (14.04)	56.00 (13.52)	0.92	$t(37)=3.88$	0.0002
Innovation (%)	17.00 (16.45)	6.00 (8.34)	0.93	$t(37)=2.69$	0.0053
Flexibility (7 point likert)	4.18 (0.81)	3.32 (0.75)	0.82	$t(37)=3.44$	0.0007
Understandability (7 point likert)	3.96 (1.03)	3.53 (0.77)	0.73	$t(37)=1.50$	0.0705
Overall Quality (%)	58.28 (13.97)	47.33 (1.177)	0.87	$t(37)=2.66$	0.0058

Hypothesis two was strongly supported ($p=0.0002$, significant with $\alpha = 0.01$) with good inter-rater reliability. As expected, data models built by expert data modellers are more complete than data models built by novice data modellers. The average score of the novice group of 56% was quite low and indicates the importance of experience in the data modelling task.

The third hypothesis was supported ($p=0.0053$, significant with $\alpha = 0.01$) with good inter-rater reliability. Data models built by expert data modellers have higher innovation than data models built by novice data modellers. The innovative concepts introduced into data models by experts may be explained by previous studies of data modelling process which showed that expert data modellers are able to categorise problem descriptions into standard abstractions and reuse generic data models from their previous experience (Batra and Davis 1992, Chaiyasut and Shanks 1994). This is in contrast to novice data modellers who developed "literal" data models from first principles.

Table 4 Summary of Correctness Results

Facet	Expert (mean/stdev)	Novice (mean/stdev)	T-value	P-value
Entity Name (%)	100.00 (0.0)	100.00 (0.0)		
Relationship Name (%)	53.89 (50.08)	29.52 (19.55)	t(37)=1.61	0.0575
Cardinality (%)	100.00 (0.0)	80.95 (9.09)	t(37)=2.68	0.0055
Mandatory / Optional (%)	86.67 (20.29)	64.29 (45.56)	t(37)=1.92	0.0310

Hypothesis four was strongly supported ($p=0.0007$, significant with $\alpha = 0.01$) with adequate inter-rater reliability. It can be seen however that the flexibility rating for experts was only just above the mid-point of the likert scale and that for novices was significantly lower than the mean ($p=0.0002$, significant with $\alpha = 0.01$) indicating that data models produced by novices have very low flexibility. Flexibility in data models is indicated by use of generalised concepts and business rules stored as data values in the data model. Both these data modelling strategies lead to the data model becoming more abstract. This result may be explained by process tracing studies which showed that expert data modellers move their solutions to higher levels of abstraction whereas novice data modellers do not (Chaiyasut and Shanks 1994).

Hypothesis five was supported ($p=0.0705$, significant with $\alpha = 0.1$) with adequate inter-rater reliability. It can be seen that the understandability rating for both experts and novices was below the mid-point of the likert scale. The novice rating in particular was significantly lower than the mean ($p=0.0052$, significant with $\alpha = 0.01$) indicating that data models built by novices have very low understandability. The low level of understandability of both expert and novice data models is consistent with empirical studies which question the useability of conceptual data models (Goldstein and Storey 1990 and Hitchman 1995). A possible explanation for the low understandability of data models built by expert data modellers is the innovative and abstract concepts they contain together with the absence of many relationship names. The poor understandability of data models built by novice data modellers may be explained by their low completeness rating and poor use of relationship names.

The sixth hypothesis was supported ($p=0.0058$, significant with $\alpha = 0.01$) with good inter-rater reliability. As expected data models built by expert data modellers are of higher overall quality than data models built by novice data modellers.

Interactions between Quality Factors

A correlation analysis was conducted to determine any interrelationships between the quality factors using Pearson product moment correlation. The results are shown in Table 5 below. The observed correlations are substantively less than 1 suggesting that the quality factors represent separate but correlated measures of quality (Gable 1996, p258). There were strong positive correlations between overall quality and understandability (.757, $p<.001$), completeness (.643, $p<.001$), and correctness (.711, $p<.001$). These three quality factors are also the three important goals in the theory-based quality framework of Krogstie *et al.* (1995).

Other interesting correlations were found between understandability and correctness (.657, $p<0.001$) and understandability and completeness (.442, $p<.001$). Krogstie *et al.* (1995) note that for a conceptual model to be readily understood it should be correct and complete. The correlation between flexibility and innovation (.628, $p<.001$) indicates that many of the innovative concepts introduced into data models by specialist data modellers may be due to the use of abstraction to make the data models more flexible. An example is the introduction of the concept *invoicable item* by one expert data modeller as a generalisation of the entity types *sale* and *repair*.

Table 5 Results of Correlation Analysis

	Correct	Complete	Understand	Innovation	Flexible	Complex	Overall
Correct	1.000						
Complete	.643 ****	1.000					
Understand	.657 ****	.442 ***	1.000				
Innovation	.395 **	.563 ****	-.146	1.000			
Flexible	.650 ****	.690 ****	.305 *	.628 ****	1.000		
Complex	.095	.491 ***	-.191	.383 *	.303 *	1.000	
Overall	.711 ****	.729 ****	.757 ****	.121	.532 ***	.167	1.000

Note: **** p < .001 *** p < .01 ** p < 0.05 * p < 0.10

A small negative correlation was found between understandability and innovation (-0.146). This indicates that the introduction of new, innovative concepts into the data model may mean that the model becomes more difficult to understand. This is an important result for specialist data modellers who use the generalisation abstraction to make their data models more flexible but at the cost of understandability. The other negative correlation, between understandability and complexity, indicates that simpler data models may be easier to understand. However, neither of these negative correlations are statistically significant.

CONCLUSION

This study found that data models built by expert data modellers are more correct, complete, innovative, flexible and better understood than those built by novices. The greatest differences between the models were found in completeness, innovation, flexibility, and in the naming of relationships. These results were related to previous empirical studies of data modelling process which offer insights into why these differences may occur.

Limitations

Empirical studies of this type have inherent limitations (Mantra 1987). The validity and generality of the results should be considered in the light of a number of factors. The participants were a group of thirty nine data modelling practitioners who had each attended a three day data modelling course and were prepared to participate in the study. The extent to which they are representative of the population of data modellers needs to be verified by further similar studies. The data modelling task was undertaken with the data modellers unable to discuss requirements with a client. However, since all participants were subjected to the same limitation the effect of this limitation is minimised. Some participants may have lacked knowledge about the domain of the case problem, although the particular problem used was selected to minimise this limitation.

Further Research

Additional research into how aspects of process and expertise lead to the differences in quality of data models is required. Outcomes from this research could be used in the design of training courses which narrow the gap between expert and novice performance. Further research is also required in how to improve understanding of conceptual data models. Krogstie *et al.* (1995) suggest a number of means to improve understanding of conceptual models, including explanation. Two approaches to explanation which may help are the use of scenarios (Carrol 1995) and argumentation-based design rationale (Buckingham Shum and Hammond 1994). The author is currently developing a methodology for designing and representing conceptual data models which synthesises Potts *et al.* (1994) inquiry based life cycle model for capturing scenarios and design rationale with the viewpoint development and integration process model of Darke and Shanks (1996). Further research will involve empirical studies of the use of the methodology in designing and representing conceptual data models.

REFERENCES

- Alexander, C. (1979) **The Timeless Way of Building**, Oxford University Press
 Avison, D. and Fitzgerald, G. (1995) **Information Systems Development: Methodologies, Techniques and Tools** (2nd edition), Oxford, Blackwell Scientific
 Barker, R. (1989) **CASE*METHOD: Entity Relationship Modelling**, Addison-Wesley

- Batra, D., Hoffer, J.A. and Bostrom, R.P. (1990) **Comparing Representations with Relational and EER Representations**, Communications of the ACM (Feb.) 33:2, p126-139
- Batra, D. and Davis, J.G. (1992) **Conceptual Data Modelling in Database Design: similarities and differences between novice and expert designers**, Int. Journal Man-Machine Studies, 37, p83-101
- Brodie, M.L., Mylopoulos, J. and Schmidt, J.W. (eds.) (1984) **On Conceptual Modelling**, Springer-Verlag, New York
- Buckingham Shum, S. and Hammond, N. (1994) **Argumentation-based Design Rationale: what use at what cost**, Int. Journal Human-Computer Studies, 40, p603-652
- Carrol, J.M. (ed.) (1995) **Scenario-Based Design**, John Wiley and Sons, New York
- Chaiyasut, P. and Shanks, G. (1994) **Conceptual Data Modelling Process: A Study of Novice and Expert Data Modellers** in T. Halpin and R. Meersman (eds.) Proc. 1st Intl. Conf. on Object-Role Modelling, Univ. of Queensland
- Darke, P. and Shanks, G. (1996) **Stakeholder Viewpoints in Requirements Definition: A Framework for Understanding Viewpoint Development Approaches**, Requirements Engineering Journal, 1:2, p88-105
- Goldstein, R.C. and Storey, V.C. (1990) **Some findings on the intuitiveness of entity-relationship constructs** in F.H. Lochovsky (ed.) Entity-Relationship Approach to Database Design and Querying, Elsevier Science, Amsterdam
- Graziano, A. M. and Raulin, M. L. (1989) **Research Methods: A Process of Enquiry**, Harper and Row
- Hitchman, S. (1995) **Practitioner Perceptions on the Use of Some Concepts in the Entity-Relationship Model**, European Journal of Information Systems, 4, 31-40
- Hoffer, J.A. (1982) **An empirical investigation into individual differences in database models** in Proc. 3rd Intl. Conf. on Information Systems, (December) p153-168
- Juhn, S. and Naumann, J. (1985) **The effectiveness of data representation on user validation** in Proc. 6th Intl. Conf. on Information Systems, Indianapolis, p212-226
- Jarvenpaa and Machesky (1989) **Data Analysis and Learning: An Experimental Study of Data Modelling Tools**, Intl. Journal Man-Machine Studies, 31, 367-391
- Kaposi, A. and Kitchenham, B. (1987) **The Architecture of System Quality**, Software Engineering Journal, 2:8
- Kim, Y-G. and March, S.T. (1995) **Comparing Data Modelling Formalisms**, Communications of the ACM, (June), 38:6, p103-113
- Konsynski, B.R. (1979) **Database Driven Systems**, University of Arizona
- Krogstie, J., Lindland, O.I. and Sindre, G. (1995) **Towards a Deeper Understanding of Quality in Requirements Engineering** Proc. CAISE
- Leviton, A. and Redman T. (1995) **Quality Dimensions of a Conceptual View**, Information Processing and Management, 31:1, p81-88
- Lindland, O., Sindre, G. and Solvberg, A. (1994) **Understanding Quality in Conceptual Modelling**, IEEE Software, (March) 42-49
- Maiden, N. and Sutcliffe, A. (1992) **Analysing the Novice Analyst: Cognitive Models in Software Engineering**, Intl. Journal Man-Machine Studies, 367, pp 719-740
- Mantra, R.W. (1987) **Data Flow and Data Structure Modelling for Database Requirements Determination: A Comparative Review**, MIS Quarterly, (December) p531-545
- Martin, J. (1987) **Recommended Diagramming Standards for Analysts and Programmers**, Prentice-Hall, New Jersey
- Moody, D. and Shanks, G. (1994) **What Makes a Good Data Model? Evaluating the Quality of Entity Relationship Models** in P. Loucopolis (eds.) Proc. Thirteenth International Entity Relationship Conference, Manchester, England
- Moody, D., Simson, G., Shanks, G., Venable, J. and Olson, N. (1995) **Stakeholder Viewpoints in Conceptual Data Modelling**, Proc. 6th Australasian Conference on Information Systems, Curtin University of Technology
- Periasamy, K.P. (1994) **Development and Usage of Information Architecture**, PhD Dissertation, University of Oxford
- Pervan, G.P. and Klass, D.J. (1992) **The Use and Misuse of Statistical Methods in Information Systems Research** in R.D. Galliers (ed.) Information Systems Research: Issues, Methods and Practical Guidelines, Blackwell, Oxford p208-229
- Potts, C., Takahashi, K. and Anton, A.I. (1994) **Inquiry-Based Requirements Analysis**, IEEE Software, (March) p21-32

- Ridjanovic, D. (1986) **Comparing quality of data representations produced by non-experts using logical data structures and relational data models**, PhD Dissertation, University of Minneapolis
- Shanks, G., Simsion, G. and Rembach, M. (1993) **The Role of Experience in Conceptual Data Modelling**, Proc. 4th Australian Conf. on Information Systems, University of Queensland
- Shanks, G. (1997) **The Challenges of Strategic Data Planning in Practice: An Interpretive Case Study**, The Journal of Strategic Information Systems, forthcoming
- Shoval, P. and Even- Chaime, M.. (1987) **Database schema design: An experimental comparison between normalisation and information analysis**, Database, 18:3 (Spring) p30-39
- Simsion, G.C. (1994) **Data Modelling Essentials**, Von Nostrand Reinhold
- Van Fliet, J.C. (1993) **Software Engineering: Principals and Practice**, John Wiley and Sons, Chichester, England