COMPLEXITY EFFECTS ON END USER UNDERSTANDING OF DATA MODELS: AN EXPERIMENTAL COMPARISON OF LARGE DATA MODEL REPRESENTATION METHODS

Daniel L. Moody

Department of Computer and Information Science
Norwegian University of Science and Technology (NTNU)
Trondheim, Norway N-7491
email: dmoody@idi.ntnu.no

School of Business Systems,
Monash University
Melbourne, Australia 3800
email: dmoody@infotech.monash.edu.au

ABSTRACT

This paper describes a laboratory experiment which evaluates the effectiveness of different representation methods for end user understanding of large data models. Data model understanding is evaluated in terms of:

- *Comprehension performance: the ability to answer questions about the data model*
- *Verification performance: the ability to identify discrepancies between the data model and a set of user requirements in textual form.*

This is the first empirical comparison of large data model representation techniques that has been conducted in over two decades of research in this area. The results suggest that there are significant complexity effects on end user understanding of data models. By reducing a data model to "chunks" of manageable size, both comprehension and verification performance can be significantly improved. This finding has implications for other graphical notations used in IS development.

1. INTRODUCTION

End User Understanding of Data Models

The Entity Relationship (ER) Model (Chen, 1976) is recognised world wide as the standard technique for data modelling in practice, and has been used to design database schemas for over two decades (Thalheim, 2000). One of the major quoted advantages of ER modelling as an analysis technique is its ability to communicate with end users. According to the literature, the ER Model is:

- Simple and easily understood by non-specialists (Konsynski, 1979);
- Highly intuitive and provides a very natural way of representing a user's information requirements (Brodie et al, 1984);
- Suitable for computer-naïve end-users (Berman, 1986).

However empirical studies show that in practice data models are poorly understood by users, and in most cases are not developed with direct user involvement (Hitchman, 1995). Experimental studies have also shown that comprehension of data models is very poor and that a large percentage of data model components are either not seen or not understood (Nordbotten and Crosby, 1999). While data modelling has proven very effective as a method for database design (as evidenced by its popularity in practice), it has been far less effective for communication with users (Goldstein and Storey, 1990; Moody, 1996; Shanks, 1997). The problem is summarised by Hitchman (1995):

> "Information from the survey gives very strong evidence to support the assertions that data modelling is poorly understood by analysts and especially by clients. This seems to contradict the widely held academic proposition that data models are easy to build and understand…"

Effects of Complexity on Data Model Understanding

One of the most serious practical limitations of the ER Model is its inability to cope with complexity (Feldman and Miller, 1986; Gilberg, 1986; Simsion, 1989; Teory et al, 1989; Gandhi et al, 1994; Akoka and Comyn-Wattiau, 1996; Allworth, 1996; Kimball, 1996; Allworth, 1999). The ER model lacks explicit abstraction mechanisms for managing the size and complexity of real world data models (Weber, 1997). With large numbers of entities, complexity quickly becomes overwhelming and as a result, data models become very difficult for people, particularly non-technical users, to understand (Feldman and Miller, 1986; Moody, 1991; Kimball, 1996).

The theoretical explanation for why people have difficulties understanding large data models is because of the limitations of human information processing—channel capacity or "cognitive bandwidth". Psychological studies show that due to limits on short-term memory, humans have a strictly limited capacity for processing information—this is estimated to be "seven, plus or minus two" concepts at a time (Miller, 1956; Newell and Simon, 1972). If the amount of information received exceeds the limits of short term memory, information overload ensues and comprehension degrades rapidly (Lipowski, 1975). Both field and experimental studies of information overload show that it results in physiological stress, psychological discomfort and reduced performance on complex tasks.

Surveys of practice show that application data models consist of an average of 95 entities, while enterprise data models consist of an average of 536 entities (Maier, 1996). Clearly, models of this size exceed human cognitive capacity many times over—about 14 times for application data models and 77 times for enterprise data models. This provides a possible explanation for why ER models are so poorly understood in practice (Goldstein and Storey, 1990; Hitchman, 1995).

Levelled Data Models

A previous paper (Moody, 1997) defined a method for representing large data models based on the organisation of a street directory. A Levelled Data Model consists of the following components (Figure 1):

- A high level diagram, called the *Context Data Model*, provides an overview of the model and how it is divided into subject areas. This corresponds to the key map in a street directory and is shown in pictorial form.
- A set of named *Subject Area Data Models* show a subset of the data model (a single subject area) in full detail. These correspond to detail maps in a street directory. *Foreign entities* are used to show cross-references between subject areas—these correspond to inter-map references in a street directory.
- An *Entity Index* is used to help locate individual entities within each subject area.

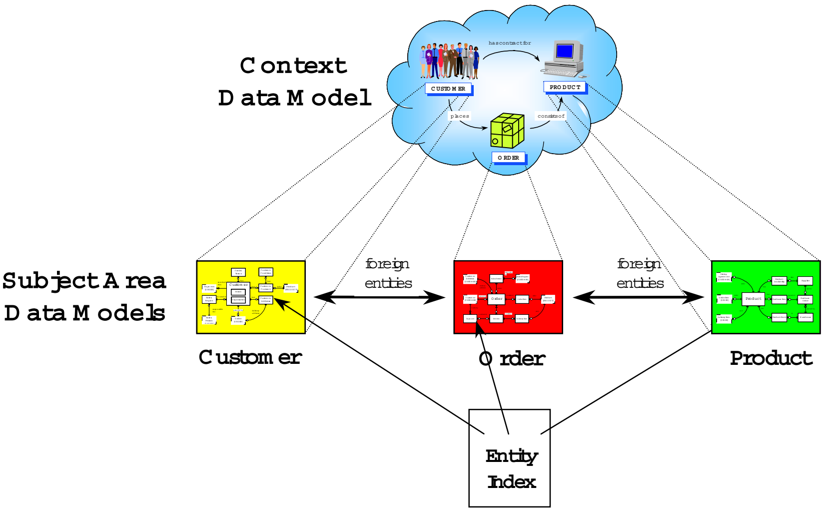

Figure 1. Levelled Data Model Architecture

The model may be organised into any number of levels, depending on the size of the underlying data model, resulting in a hierarchy of models at increasing levels of detail.

Research Questions

This paper describes a laboratory experiment which evaluates the effectiveness of the proposed method (Levelled Data Models) for improving user understanding of large data models compared to the standard ER model and methods previously proposed in the literature. The two broad research questions addressed by this experiment are:

1. Are large data models represented using the proposed method more easily understood by end users than models represented in standard ER form?
2. Are large data models represented using the proposed method more easily understood by end users than models represented using methods previously proposed in the literature?

2. PREVIOUS RESEARCH

Review of Large Data Model Representation Methods

While the representation of large data models is an important issue in practice, it is a problem that has attracted relatively little research attention. A number of methods have been proposed to address this issue (e.g. Martin and McClure, 1985; Feldman and Miller, 1986; Gilberg, 1986; Simsion, 1989; Teory et al, 1989; Gandhi et al, 1994; Allworth, 1996; 1999), but none have been widely adopted in practice. The greatest weakness in the existing literature is the lack of empirical validation of these methods. So far, there has been no systematic empirical research into the effectiveness of these methods in practice. The authors of the methods argue that their approaches are effective but in most cases, no empirical evidence is provided. Most evidence of successful use of these methods is anecdotal and in many cases reports the direct experience of the author (Shanks, 1996).

Review Of Experimental Studies Of Data Model Understanding

While there have been no previous experimental studies of large data model representational methods, there have been a number of experimental studies of data model understanding (Juhn and Naumann, 1985; Leitheiser, 1988; Palvia et al, 1992; Shoval and Frumerman, 1994; Hardgrave and Dalal, 1995; Kim and March, 1995; Nordbotten and Crosby, 1999; Bodart et al, 2001). However there are a number of methodological weaknesses in these studies, particularly with respect to external validity.

Size of Models Used

A criticism of all previous experimental studies is that the models used are not of realistic size and complexity. The largest model used in any of these studies was 15 entities, which is significantly smaller than the average size application data model (≈ 95 entities). This is an external validity issue: are the results obtained generalisable to models of real world size and complexity?

Graphical Notation Used

Another criticism of previous experimental studies is that the notations used are not representative of those used in practice. The Extended Entity Relationship (EER) model (Elmasri and Navathe, 1994) is an "academic form" of the ER model which is based on Chen's original notation (Chen, 1976). The EER notation is popular in textbooks, university courses and research studies, but is rarely used in practice (Hitchman, 1995). The EER notation has been used in almost all experimental studies of data modelling understanding, which brings into question the generalisability of their results.

Experimental Tasks

To maximise external validity, it is desirable that the experimental task simulate as closely as possible what people would be required to do in the real world (Baker, 1998). A criticism of most of the previous experimental studies of data model understanding is that the tasks used to test user comprehension are not representative of tasks that users are required to perform in practice. Only one of the studies (Kim and March, 1995) uses verification of the model against a specification, even though this is what users are generally required to do in the real world. Tasks such as recall, comprehension, verbal description and problem solving would rarely be required of users in real life.

3. RESEARCH DESIGN

Research Method Selection

There are a wide variety of research methods which may be used in conducting IS research (Galliers, 1991; Nunamaker et al, 1991; Galliers, 1992; Shanks et al, 1993; Baskerville and Wood-Harper, 1996; Wynekoop and Russo, 1997). Different research methods are appropriate in different situations, depending on the research question and the stage of knowledge in the area being studied (Galliers, 1991; Shanks et al, 1993; Wynekoop and Russo, 1997). In general, a combination of research methods may be most effective in achieving a particular research objective (Jick, 1979; Kaplan and Duchon, 1988; Fitzgerald, 1991; Lee, 1991; Galliers, 1992; Wynekoop and Russo, 1997; Neuman, 2000). For example, when a subject area is not well understood, qualitative methods may be used to build theory and testable hypotheses. Theory may then be tested using quantitative methods such as surveys and experiments.

Prior to this study, the proposed method had been extensively tested in practice using an *action research* approach (Moody, 2001; Moody, 2002). It was applied in eight different organisations as part of an ongoing action research programme. The method was refined significantly as a result of use in practice, and reached a point where it was a stable and mature approach (as evidenced by the lack of change from one action research cycle to the next). While action research was an appropriate research method when the method was in its developmental phases, it was clearly less suitable in evaluating the method once it had become stable—this is similar to the difference between theory building (exploratory research) and theory testing (evaluation research). A controlled experiment provides the most effective way to evaluate the effectiveness of the proposed method because:

- It allows direct comparisons to be made between different methods under controlled conditions through manipulation of experimental treatments.
- It enables the method to be evaluated using objective and quantitative data.
- It enables the method to be evaluated using independent participants.
- It is possible to establish that the attainment of the objectives was attributable to the use of the method, by factoring out all other variables which may have contributed to the outcomes.

Mixing qualitative and quantitative research methods is called *triangulation of method* (Neuman, 2000). The two types of methods have different, complementary strengths and when used together can lead to a more comprehensive understanding of a phenomenon (Jick, 1979; Kaplan and Duchon, 1988).

Identifying Methods for Comparison

To make the experiment manageable, it was decided to limit the number of methods evaluated as part of Research Question 2 to the two leading methods proposed in the literature. The selection of methods was based on the following criteria, which represent a balance between rigour and relevance:

- Academic credibility (*rigour*): publication in a refereed academic journals
- Practical credibility (*relevance*): evidence of successful use in practice

Only two methods satisfied both of these criteria: Clustered Entity Models (Feldman and Miller, 1986) and Structured Data Models (Simsion, 1989), so these were the methods that were evaluated. These methods also represent the two predominant paradigms for clustering models: *aggregation* and *generalisation*.

Experimental Design

A four group, post-test only design was used, with one active between-groups factor (representation method). The experimental design is summarised in Figure 2.

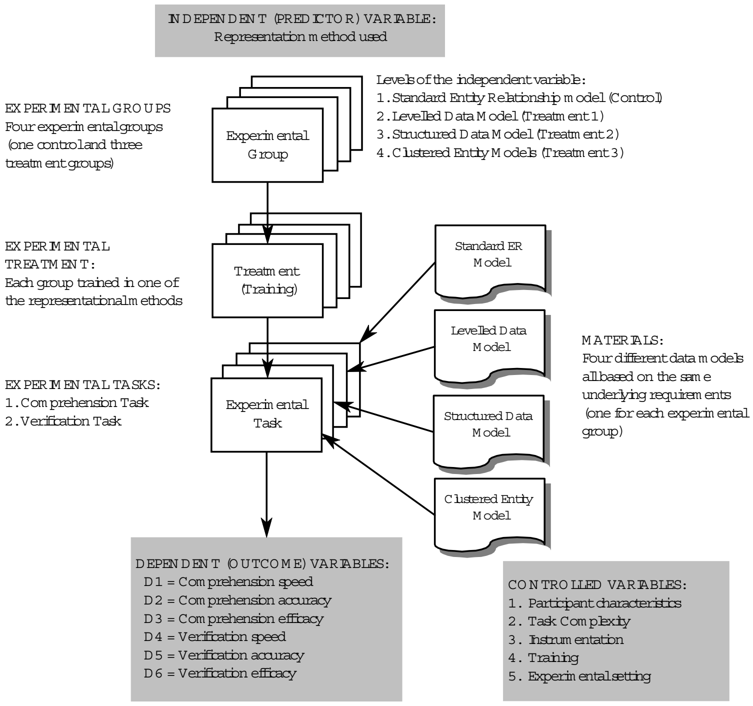

Figure 2. Experimental Design

The experimental groups consisted of a control group (standard ER model) and three treatment groups (the proposed method and the two leading "competitor" methods from the literature). Participants were trained in the conventions of one of the methods, and then given an example data model represented using this method. They were then given a set of questions to answer about the model (*comprehension task*) and a description of user requirements which they were asked to verify the model against (*verification task*).

Independent Variable

The independent variable is the method used to represent the experimental data model (Representation Method). The independent variable has four levels, corresponding to the different representation methods being evaluated:

- Standard ER representation
- Levelled Data Models (the proposed method)
- Clustered Entity Models
- Structured Data Models

Dependent Variables

User validation of data models consists of two separate cognitive processes: comprehension and verification (Kim and March, 1995). Users must first *comprehend* or understand the meaning of the model, then they must *verify* the model by identifying any discrepancies between the model and their (tacit) knowledge of their requirements. According to Kim and March (1995), comprehension performance reflects *syntactic understanding*: the person's competence in understanding the constructs of the modelling formalism, while verification performance reflects *semantic understanding*: the person's ability to apply that understanding. In defining measures of understanding, we therefore distinguish between two *types* of performance:

- Comprehension performance: the ability to answer questions about a data model
- Verification performance: the ability to identify discrepancies between a data model and a given set of user requirements.

We also distinguish between three *dimensions* of performance:

- Efficiency: the effort required to understand a model (this requires measuring task *inputs).*
- Effectiveness: how well the data model is understood (this requires measuring task *outputs*)
- Efficacy: the combination of efficiency and effectiveness (this requires measuring the ratio of outputs to inputs)

In this experiment, we define six dependent variables, which cover all aspects of understanding performance.

Table 1. Classification of Dependent Variables

	EFFICIENCY	EFFECTIVENESS	EFFICACY
COMPREHENSION PERFORMANCE	D1: Comprehension Efficiency (speed)	D2: Comprehension Effectiveness (accuracy)	D3: Comprehension Efficacy (accuracy/speed)
VERIFICATION PERFORMANCE	D4: Verification Efficiency (speed)	D5: Verification Effectiveness (accuracy)	D6: Verification Efficacy (accuracy/speed)

Hypotheses

The research questions defined in Section 1 are broken down into several *hypotheses*, each relating to a particular combination of levels of the independent variable and one of the dependent variables. This results in 12 separate hypotheses. All of the hypotheses involve comparisons between the proposed method and the standard ER model (Research Question 1) or between the proposed method and methods previously proposed in the literature (Research Question 2).

Research Question 1

Comprehension performance:

- H1: Participants using the Levelled Data Model will perform the comprehension task faster than those using the standard ER Model
- H2: Participants using the Levelled Data Model will perform the comprehension task more accurately than those using the standard ER Model
- H3: Participants using the Levelled Data Model will perform the comprehension task faster and more accurately than those using the standard ER Model

Verification performance:

- H4: Participants using the Levelled Data Model will perform the verification task faster than those using the standard ER Model
- H5: Participants using the Levelled Data Model will perform the verification task more accurately than those using the standard ER Model
- H6: Participants using the Levelled Data Model will perform the verification task faster and more accurately than those using the standard ER Model

Rationale for H1-H6. The rationale for these hypotheses is that the standard ER Model does not provide any explicit abstraction mechanisms for dealing with complexity (Moody, 1997; Weber, 1997). This will result in a state of information overload for participants in performing the comprehension and verification tasks. As a result, both speed and accuracy of performance will be reduced (Lipowski, 1975; Baddeley, 1999). The Levelled Data Model decomposes the data model into parts of cognitively manageable size, which will reduce the cognitive complexity of the task and therefore improve comprehension and verification performance.

Research Question 2

Comprehension performance:

- H7: Participants using the Levelled Data Model will perform the comprehension task faster than participants using models represented using the competitor methods
- H8: Participants using the Levelled Data Model will perform the comprehension task more accurately than participants using models represented using the competitor methods
- H9: Participants using the Levelled Data Model will perform the comprehension task faster and more accurately than participants using models represented using the competitor methods

Verification performance:

- H10: Participants using the Levelled Data Model will perform the verification task faster than participants using models represented using the competitor methods
- H11: Participants using the Levelled Data Model will perform the verification task more accurately than participants using models represented using the competitor methods
- H12: Participants using the Levelled Data Model will perform the verification task faster and more accurately than participants using models represented using the competitor methods

Rationale for H7-H12. The Structured Data Model representation (Treatment Group 2) is expected to be less efficient and effective than the Levelled Data Model because:

- It uses generalisation to structure the data model, which is less natural from a human cognition viewpoint than aggregation (Eysenck and Keane, 2000). The concept of generalisation is not well understood even by experienced data modellers (Hitchman, 1995).
- Important requirements information is lost as a result of the transformation, which makes it difficult or impossible to answer certain questions.
- The method does not reduce the model to parts of manageable size: in the experimental data model, all diagrams exceed 7 ± 2 concepts, which means that information overload will still be a problem.

The Clustered Entity Model representation (Treatment Group 3) is expected to be less efficient and effective than the Levelled Data Model because:

- The use of relationships between objects at different refinement levels will lead to problems in interpretation of the model (Klir, 1985).
- The number of levels and diagrams will make it more difficult to find relevant information.
- The problems of construct overload and construct ambiguity will lead to difficulties in interpreting the model (Weber, 1997).

Participants

There were 60 participants in the experiment, all of whom were first year Accounting students at the University of Melbourne. Subjects had no prior experience in the use of data modelling techniques (which was a condition of selection for the experiment), so were considered as proxies for naïve users. All participated voluntarily in the experiment and were paid $25 on completion of the experiment. Subjects were randomly assigned to experimental groups.

Materials

Four different data models were prepared for use in the experiment. All were based on the same underlying data model, but were represented using one of the methods defined by the independent variable. Different data models were given to different experimental groups. The underlying data model used was a data model for a customer management and billing system. This consisted of 98 entities (109 including subtypes), so was close to the average size for an application data model. The model consisted 480 attributes.

Experimental Treatment

Each experimental group was given a thirty minute training session in one of the representation methods plus a five minute description of the domain represented by the experimental data model. To ensure the provision of equivalent training for the four experimental groups, the same example data model and similar instructional materials were used for all groups. Each subject was also given a one page summary of the conventions and rules of the method to refer to during the experimental tasks. While the training time allowed may seem quite short, it is realistic, in that end users would be allowed a similar amount of training time in real life.

Experimental Tasks

Comprehension Task

A set of 25 true/false questions was developed to test comprehension performance. Participants were required to answer these questions in the comprehension task.

Verification Task

A one page textual description of requirements was developed to test verification performance. Participants were required to identify discrepancies between the stated requirements and the data model in the verification task. There were 15 discrepancies between the data model and the set of requirements.

4. RESULTS AND DISCUSSION

A one-way analysis of variance (ANOVA) was used to analyse differences between experimental groups on all dependent variables. Planned comparisons were conducted using predefined contrasts, while *post hoc* comparisons were conducted using *Tukey's Honestly Significant Difference (HSD) test*.

Comprehension Efficiency

Table 2 shows the results for Comprehension Time for each experimental group. Surprisingly, subjects using the Levelled Data Model took the longest to complete the comprehension task, which was the opposite of what was expected. Two of the comparisons between groups were found to be statistically significant, but both in the *reverse direction* to that predicted—subjects using the standard ER model ($\alpha < .01$) and the Structured Data Model ($\alpha < .05$) took less time to complete the task than subjects using the Levelled Data Model. This means that hypotheses H1 and H8 were *not* supported.

Table 2. Comprehension Time Statistics

EXPERIMENTAL GROUP	MEAN (μ)	STDEV (δ)
Standard ER Model	26.67	7.99
Levelled Data Model	34.00	6.02
Structured Data Model	25.82	9.24
Clustered Entity Model	30.64	6.39

Comprehension Effectiveness

Table 3 shows the results for Comprehension Accuracy for each experimental group. As predicted, participants using the Levelled Data Model performed the best, and scored 17% better than those using the standard ER model.

Table 3. Comprehension Accuracy Statistics

EXPERIMENTAL GROUP	MEAN (μ)	STDEV (δ)
Standard ER Model	69.60%	9.66%
Levelled Data Model	**81.43%**	10.48%
Structured Data Model	66.82%	10.93%
Clustered Entity Model	70.57%	12.41%

On the face of it, the mean scores indicate that participants understood the data model reasonably well—the results for all groups were above 65%. However the fact that the questions were all True/False most likely overstates comprehension performance, because participants would have scored 50% based on chance alone. When the scores are adjusted for chance (by subtracting 50% and multiplying the result by two, giving the percentage improvement on chance), the results for all except Treatment Group 2 are less than 50%. Using this measure, participants using the Levelled Data Model scored 63% better than those using the standard ER model.

Table 4. Comprehension Accuracy Scores Adjusted for Chance

EXPERIMENTAL GROUP	MEAN (μ)
Standard ER Model	39.20%
Levelled Data Model	62.86%
Structured Data Model	33.65%
Clustered Entity Model	41.14%

Subjects using the Levelled Data Model performed significantly better than all other groups ($\alpha < .05$), but there was no difference between any of the other groups. This confirms both H2 and H9.

Comprehension Efficacy

Table 5 shows the results for normalised comprehension accuracy for each experimental group.

Table 5. Normalised Comprehension Accuracy Statistics

EXPERIMENTAL GROUP	MEAN (μ)	STDEV (δ)
Standard ER Model	41.86	11.39
Levelled Data Model	37.24	9.39
Structured Data Model	**46.30**	27.29
Clustered Entity Model	35.52	8.28

Because this variable was found to non-normally distributed within two of the experimental groups, non-parametric methods were used to evaluate differences between groups. The Kruskal-Wallis H test is the nonparametric equivalent of ANOVA and was used to evaluate both planned and post hoc comparisons. No significant differences between any of the experimental groups. This means that hypotheses H3 and H10 were not supported.

From these results, it would appear that subjects did indeed make trade-offs between time taken and accuracy. The best performed group in terms of accuracy (Levelled Data Model) took the longest time, while the worst performed group in terms of accuracy (Structured Data Model) took the shortest time. A significant correlation was found between Comprehension Time and Comprehension Accuracy ($r = .364$, $\alpha = .004$**). However, while a nominal time limit was set for the task, this was not strictly enforced. Therefore, subjects had no strong need to make trade-offs between time and task performance apart from a desire to finish the experiment as quickly as possible. An alternative explanation is that Treatment Group 2 took such a short time because subjects *guessed* a significant percentage of their answers. Because the Structured Data Model representation results in loss of information from the original model, in many cases it would have been impossible to answer the comprehension questions based on the information provided. Where subjects were unable to answer the question, they may have guessed the answer, which in a True/False question gives a 50/50 chance of being correct. This would also explain the high variability in the results for Treatment Group 2.

This suggests a systematic flaw in all previous experimental studies which have evaluated user comprehension of data models, all of which have used true/false questions. Data models naturally lend themselves to questions in a true/false format e.g. Can a project have many employees? Can an employee manage many projects? However using true/false questions introduces a significant amount of measurement error, as subjects can score 50% simply by guessing. A better approach in future may be to use multiple choice questions, with an "Unable to Tell" option.

Verification Efficiency

Table 6 shows the results for verification time for each experimental group. As for the comprehension task, participants using the Structured Data Model took the shortest time. None of the comparisons were significant ($\alpha < .05$). This means that hypotheses H4 and H11 were *not* supported.

Table 6. Verification Time Statistics

EXPERIMENTAL GROUP	MEAN (μ)	STDEV (δ)
Standard ER Model	31.87	7.58
Levelled Data Model	31.71	7.47
Structured Data Model	**31.41**	8.23
Clustered Entity Model	35.21	12.68

Verification Effectiveness

Table 7 shows the results for verification accuracy for each experimental group. As expected, participants using the Levelled Data Model performed the best, and scored 59% better than those using the standard ER model.

Table 7. Verification Accuracy Statistics

EXPERIMENTAL GROUP	MEAN (μ)	STDEV (δ)
Standard ER Model	37.87%	21.37%
Levelled Data Model	**60.06%**	21.78%
Structured Data Model	14.44%	16.64%
Clustered Entity Model	31.81%	18.53%

Significance testing showed that subjects using the Levelled Data Model performed significantly better than all other experimental groups ($\alpha < .01$). This means that hypotheses H5 and H12 were strongly supported. Neither of the competitor methods improved verification performance compared to the standard ER model, and the Structured Data Model group performed significantly worse than those using the standard ER model ($\alpha < .01$).

A point worth noting about these results is the low scores for all groups on this task. Three out of the four groups were below the 50% level and one of the groups was below the 20% level. This suggests that it may be expecting too much of end users to verify the correctness of data models. The inability of users to perform this task effectively may explain the high level of requirements errors reported in practice (van Vliet, 1993).

The low performance levels on this task also confirms the earlier suspicion that subjects resorted to guessing in the comprehension task, as the performance levels are significantly below those achieved on the comprehension test. In the comprehension test, subjects were rewarded for guessing (they had a 50% chance of getting the right answer), while in this task, they were penalised for incorrect answers. Overall, subjects performed less than half as well on this task (35%) as on the comprehension task (71.8%).

Verification Efficacy

Table 8 shows the results for normalised verification accuracy for each experimental group.

Table 8. Normalised Verification Accuracy Statistics

EXPERIMENTAL GROUP	MEAN (μ)	STDEV (δ)
Standard ER Model	7.82	2.92
Levelled Data Model	**12.56**	5.51
Structured Data Model	3.32	3.02
Clustered Entity Model	7.11	5.05

Subjects using the Levelled Data Model performed significantly better than all other experimental groups ($\alpha < .01$), which means that hypotheses H6 and H13 were strongly supported. The Structured Data Model group performed significantly worse than those using the standard ER model ($\alpha < .05$). The results for normalised verification accuracy were very similar to the results for verification accuracy, which suggests there were no trade-offs between time and accuracy on this task. This confirms the earlier suspicion that the apparent tradeoffs on the comprehension task were due to guessing, as in this task, guesswork was taken out of the equation.

5. CONCLUSION

Summary of Findings

Of the 12 hypotheses originally proposed, 6 were supported, and 6 were not supported, with a reverse finding in one case (H1).

Effectiveness

Of the four hypotheses relating to the effectiveness of the method (H2, H5, H9, H12), all were supported. The Levelled Data Model was found to significantly improve both comprehension and verification accuracy compared to the standard ER model and both competitor methods. The results show that use of the method improves end user comprehension and verification compared to the standard ER model by more than 50%,

Efficiency

None of the hypotheses relating to the efficiency of the method (H1, H4, H8, H11) were supported (in fact the reverse result was found for H1). However this is less important from a practical viewpoint—simply saving time in the validation process would be unlikely to be seen by practitioners as sufficient justification for adopting a new method. The *cost of time* is almost insignificant compared to the *cost of errors* in the validation process.

Efficacy

Mixed results were found for the efficacy of the proposed method. Of the four hypotheses relating to efficacy (H3, H6, H10, H13), two were supported and two were not supported. In the comprehension task, time taken and accuracy seemed to be inversely related, while on the verification task, all groups took about the same time. Most subjects were able to complete the experimental tasks within the time allowed, so there seemed little need for subjects to make tradeoffs between time and accuracy. Unless there is a strict time limit in the experimental task, this is probably not a useful construct to measure.

Surprisingly, post hoc testing between groups showed that neither of the competitor methods were found to be superior to the standard ER model on any of the dependent variables—the Structured Data Model was found to be inferior to the standard ER model on three of the dependent variables. This may explain why neither of these methods have been widely adopted in practice—this may show good judgement on the part of data modelling practitioners!

Practical Significance

This research has important implications for the practice of data modelling. The ER model was originally proposed as a means of communicating with non-technical users. However while it has proven very effective as a technique for database design, it has been far less effective for communication with users (Goldstein and Storey, 1990; Moody, 1996; Nordbotten and Crosby, 1999).

This paper has argued that one of the major reasons for the difficulty people have in understanding data models is the limitations of human cognitive capacity (cognitive bandwidth or channel capacity). The results of this experiment provide strong evidence for information overload effects on data model understanding, which may provide an explanation for why data models are so poorly understood in practice. The results show that the Levelled Data Model notation significantly improves user understanding of data models compared to the standard ER model (62% improvement in comprehension performance and 59% improvement in verification performance). By reducing a data model to "chunks" of manageable size, which match the channel capacity of the human mind, complexity is reduced and understanding is improved.

To my knowledge, this is the first extension to the ER Model that has been empirically shown to improve user understanding of data models compared to the standard ER Model in over 25 years of research in this area. Literally hundreds of extensions to the ER Model have been proposed in the literature, but very few of these have been tested empirically. Theoretical justification is limited as a form of justification because methods have no "truth" value—the validity of a method is an empirical rather than a theoretical question (Rescher, 1977; Ivari, 1986).

Theoretical Significance

This is the first experimental evaluation of large data model representational techniques that has been conducted in almost two decades of research in this area. This experiment provides empirical evidence about the comparative effectiveness of different representation methods for user understanding compared to the standard ER model.

The findings on the complexity effects on understanding of data models have implications for all types of graphical notations used in IS design. Practically all systems development techniques use graphical representations (Nordbotten and Crosby, 1999), and it is likely that information overload is an issue in all such techniques. The principles used in Levelled Data Models to reduce the complexity of large data models could be applied to other analysis and design techniques.

Strengths and Weaknesses of the Research

Internal Validity

The following variables were controlled as part of this experiment:

- Participant Characteristics: these were controlled by random assignment to experimental groups.
- Task Complexity: the same underlying data model was used by all groups.
- Instrumentation: the same instruments were used to measure dependent variables for all experimental groups.
- Training: the same amount of training was provided to each experimental group.

- Experimental Setting: the location, time of day, time of year, the experimenter and instructions given to subjects were consistent across experimental groups.

External Validity

The artificiality of the laboratory is one of the major disadvantages of the experimental method (Cooper and Schindler, 1998). To maximise external validity, experimental conditions should be made as similar as possible to conditions to which results are to be generalised. The following strategies were used to increase the external validity of the experiment:

- Using a model of realistic size and complexity. All previous experimental studies have used example models that were extremely simple, even trivial in size (Shoval, 1997). The data model used in this experiment consisted of 98 entities, which is around the mean size for an application data model, and is more than eight times the size of any data model previously used.
- Using the representation of the ER model that is commonly used in practice, and which end users would be most likely to deal with. All previous experimental studies have used an "academic" ER notation (the EER model) that is rarely used in practice.
- Using a "natural" data model taken from a real world application project rather than one developed for the purposes of the experiment.
- Using verification to test understanding of the model, which most closely approximates the task that users would be required to perform in practice. Only one previous experimental study has used verification to test understanding of data models (Kim and March, 1995).

Another possible threat to external validity was the use of students as experimental subjects. Generalisability is a significant problem in most laboratory experiments involving students, and has been identified as a major issue in IS research in terms of its relevance to practice (Keen, 1991a; Galliers, 1994; Moody, 2000). However because the selection criteria for subjects was that they had no previous exposure to ER Modelling, they can be considered as reasonable proxies for naïve users.

REFERENCES

1. AKOKA, J. and COMYN-WATTIAU, I. (1996): "Entity Relationship And Object Oriented Model Automatic Clustering", *Data And Knowledge Engineering*, **20**, 1,
2. ALLWORTH, S. (1996): "Using Classification Structures To Develop And Structure Generic Industry Models", *Proceedings Of The First Australian Data Management Conference*, D.L. Moody, (Ed.), Melbourne, Australia, Australian Data Management Association (DAMA).
3. ALLWORTH, S. (1999): "Classification Structures Encourage the Growth of Generic Industry Models", *Proceedings of the Eighteenth International Conference on Conceptual Modelling (Industrial Track)*, D.L. Moody, (Ed.), Paris, France, Springer.
4. BABBIE, E.R. (1998): *The Practice of Social Research*, Wadsworth Publishing, Belmont, California.
5. BADDELEY, A.D. (1999): *Essentials of Human Memory*, Psychology Press, Hove [England].
6. BAKER, T.L. (1998): *Doing Social Research*, McGraw-Hill.
7. BASKERVILLE, R.L. and WOOD-HARPER, T. (1996): "A Critical Perspective on Action Research as a Method for Information Systems Research", *Journal of Information Technology*, **3**, 11, pp. 235-246,
8. BODART, F., PATEL, A., SIM, M. and WEBER, R. (2001): "Should the Optional Property Construct be used in Conceptual Modelling: A Theory and Three Empirical Tests", *Information Systems Research*, **(forthcoming)**,
9. CHEN, P.P. (1976): "The Entity Relationship Model: Towards An Integrated View Of Data", *ACM Transactions On Database Systems*, **1**, 1, pp. 9-36, March.
10. ELMASRI, R. and NAVATHE, S.B. (1994): *Fundamentals of Database Systems*, Benjamin Cummings, Redwood City, California.
11. EYSENCK, M.W. and KEANE, M.T. (2000): *Cognitive Psychology: A Student's Handbook*, Lawrence Erlbaum, Hove.
12. FELDMAN, P. and MILLER, D. (1986): "Entity Model Clustering: Structuring A Data Model By Abstraction", *The Computer Journal*, **29**, 4,
13. FITZGERALD, G. (1991): "Validating New Information Systems Techniques: A Retrospective Analysis", *Information Systems Research: Contemporary Approaches And Emergent Traditions*, H.E. Nissen, H.K. Klein and R. Hirschheim, (Eds.), North-Holland.

14. GALLIERS, R.D. (1991): “Choosing Information Systems Research Approaches”, *Information Systems Research: Contemporary Approaches And Emergent Traditions*, H.E. Nissen, H.K. Klein and R. Hirschheim, (Eds.), North-Holland.
15. GALLIERS, R.D. (1992): *Information Systems Research: Issues, Methods and Practical Guidelines*, Blackwell Scientific Publications.
16. GANDHI, M., ROBERTSON, E.L. and VAN GUCHT, D. (1994): “Levelled Entity Relationship Models”, *Proceedings Of The Thirteenth International Conference On The Entity Relationship Approach*, P. Loucopolous, (Ed.), Manchester.
17. GILBERG, R.F. (1986): “A Schema Methodology For Large Entity-Relationship Diagrams”, *Proceedings Of Fourth International Conference On The Entity Relationship Approach*, P.P. Chen, (Ed.).
18. GOLDSTEIN, R.C. and STOREY, V.C. (1990): “Some Findings On The Intuitiveness Of Entity Relationship Constructs”, *Entity Relationship Approach To Database Design And Querying*, F.H. Lochovsky, (Ed.), Amsterdam, Elsevier Science.
19. HARDGRAVE, B.C. and DALAL, N.P. (1995): “Comparing Object Oriented and Extended Entity Relationship Models”, *Journal Of Database Management*, **6**, 3,
20. HITCHMAN, S. (1995): “Practitioner Perceptions On The Use Of Some Semantic Concepts In The Entity Relationship Model”, *European Journal Of Information Systems*,
21. IVARI, J. (1986): “Dimensions Of Information Systems Design: A Framework For A Long Range Research Program”, *Information Systems Journal*, June,
22. JICK, T.D. (1979): “Mixing Qualitative and Quantitative Methods: Triangulation in Action”, *Administrative Science Quarterly*, **24**,
23. JUHN, S. and NAUMANN, J. (1985): “The Effectiveness of Data Representation Characteristics on User Validation”, *Proceedings of the Sixth International Conference on Information Systems*.
24. KAPLAN, B. and DUCHON, D. (1988): “Combining Qualitative and Quantitative Methods in Information Systems Research: A Case Study”, *MIS Quarterly*, December.
25. KIM, Y.-G. and MARCH, S.T. (1995): “Comparing Data Modelling Formalisms”, *Communications of the ACM*, **38**, 6,
26. KIMBALL, R. (1996): *The Data Warehouse Toolkit*, J. Wiley & Sons, New York.
27. KLIR, G.J. (1985): *Architecture of Systems Problem Solving*, Plenum Press, New York.
28. LEE, A. (1991): “Integrating Positivist and Interpretivist Approaches to Organisational Research”, *Organisational Science*, **2**, 4, pp. 342-365, November.
29. LEITHEISER, R.L. (1988): *An Examination Of The Effects Of Alternative Schema Descriptions On The Understanding Of Database Structure And The Use Of A Query Language*, University of Minnesota Ph.D Dissertation.
30. LIPOWSKI, Z.J. (1975): “Sensory And Information Inputs Overload: Behavioural Effects”, *Comprehensive Psychiatry*, **16**, 3, May/June.
31. MAIER, R. (1996): “Benefits And Quality Of Data Modelling-Results Of An Empirical Analysis”, *Proceedings Of The Fifteenth International Conference On The Entity Relationship Approach*, B. Thalheim, (Ed.), Cottbus, Germany, Elselvier.
32. MARTIN, J. and MCCLURE, C.L. (1985): *Diagramming Techniques for Analysts and Programmers*, Prentice-Hall, Englewood Cliffs, N.J.
33. MOODY, D.L. (1991): “A Practical Methodology For The Representation Of Large Data Models”, *Proceedings Of The Australian Database And Information Systems Conference*, University Of N.S.W., Sydney, Australia.
34. MOODY, D.L. (1996): “Graphical Entity Relationship Models: Towards A More User Understandable Representation Of Data”, *Proceedings Of The Fourteenth International Conference On The Entity Relationship Approach*, B. Thalheim, (Ed.), Cottbus, Germany.
35. MOODY, D.L. (1997): “A Multi-Level Architecture For Representing Enterprise Data Models”, *Proceedings Of The Sixteenth International Conference On Conceptual Modelling (ER'97)*, D.W. Embley and R.C. Goldstein, (Eds.), Los Angeles.
36. MOODY, D.L. (2001): *Dealing with Complexity: A Practical Method for Representing Large Entity Relationship Models (PhD Thesis)*, Department Of Information Systems, University of Melbourne, Melbourne, Australia.
37. MOODY, D.L. (2002): “Validation of a Method for Representing Large Entity Relationship Models: An Action Research Study”, *Proceedings of the Tenth European Conference on Information Systems (ECIS'2002)*, Gdansk, Poland.
38. NEUMAN, W.L. (2000): *Social Research Methods - Qualitative and Quantitative Approaches (4th edition)*, Allyn and Bacon, Needham Heights, MA.

39. NORDBOTTEN, J.C. and CROSBY, M.E. (1999): "The Effect of Graphic Style on Data Model Interpretation", *Information Systems Journal*, **9**,
40. NUNAMAKER, J., CHEN, M. and PURDIN, T.D.M. (1991): "Systems Development in Information Systems Research", *Journal of Management Information Systems*, **7**, 3, pp. 89-106, Winter.
41. PALVIA, T., LIO, C. and TO, P. (1992): "The Impact of Conceptual Data Models on End User Performance", *Journal Of Database Management*, **3**, 4,
42. RESCHER, N. (1977): *Methodological Pragmatism: Systems-Theoretic Approach to the Theory of Knowledge*, Basil Blackwell, Oxford.
43. SHANKS, G., ROUSE, A. and ARNOTT, D. (1993): "A Review of Approaches to Research and Scholarship in Information Systems", *Proceedings of the 4th Australian Conference on Information Systems*, Brisbane.
44. SHANKS, G.G. (1996): *Building And Using Corporate Data Models (PhD Thesis)*, Department Of Information Systems, Monash University, Melbourne, Australia.
45. SHANKS, G.G. (1997): "The Challenges Of Strategic Data Planning In Practice: An Interpretive Case Study", *Journal Of Strategic Information Systems*, **6**,
46. SHOVAL, P. and FRUMERMAN, I. (1994): "OO and EER Schemas: A Comparison of User Comprehension", *Journal Of Database Management*, **5**, 4,
47. SIMSION, G.C. (1989): "A Structured Approach To Data Modelling", *The Australian Computer Journal*, August.
48. TEORY, T.J., WEI, G., BOLTON, D.L. and KOENIG, J.A. (1989): "ER Model Clustering As An Aid For User Communication And Documentation In Database Design", *Communications Of The ACM*, August.
49. THALHEIM, B. (2000): *Entity Relationship Modeling: Foundations of Database Technology*, Springer, Berlin ; New York.
50. VAN VLIET, J.C. (1993): *Software Engineering: Principles and Practice*, John Wiley and Sons, Chichester, England.
51. WEBER, R.A. (1997): *Ontological Foundations Of Information Systems*, Coopers And Lybrand Accounting Research Methodology Monograph No. 4, Coopers And Lybrand, Melbourne, Australia.
52. WYNEKOOP, J.L. and RUSSO, N.L. (1997): "Studying Systems Development Methodologies: An Examination Of Research Methods", *Information Systems Journal*, **7**, 1, January.