Proceedings of the 2006 OR Society Simulation Workshop
S. Robinson, S. Taylor, S. Brailsford and J.Garnett, eds.

IMPROVING THE UNDERSTANDING OF CONCEPTUAL MODELLING

Wang Wang
Roger J. Brooks
Department of Management Science
Lancaster University Management School
Lancaster
LA1 4YX
w.wang3@lancaster.ac.uk, roger.brooks@lancaster.ac.uk

ABSTRACT:
Conceptual modelling, determining what to include in the model and at what level of detail, affects model performance directly. However, it has so far received little attention in the literature mainly because conceptual modelling is viewed as more of an art than science. This paper describes two preliminary studies on conceptual modelling. In the first study, data on the time spent on the different topics during a simulation project by an expert and by novice modellers was collected and analysed. The second study was an experiment to investigate the effects of different aspects of model complexity on the ease of understanding of the model and on model build time.

Keywords: Conceptual model, Model selection, Model performance

1. INTRODUCTION

Conceptual modelling is a crucial stage of the simulation modelling process, and yet it is poorly understood. Robinson [1] defined the conceptual model as a "*non-software specific description of the simulation model that is to be developed, describing the objectives, inputs, outputs, content, assumptions and simplifications of the model*". Conceptual modelling therefore involves deciding the way in which the virtual world of the simulation model should work. Law [2] considered that for simulation projects "the most difficult aspect of a study is that of determining the appropriate level of model detail". However, little attention is devoted to conceptual modelling in most textbooks.

The advice that is provided often centres on the complexity or level of detail of the model. For example, Robinson [3] proposed that the basic rule for what to include in a model is to use the minimum components required to achieve the project's objective. In fact, "Model Simple – Think complicated" is one of Pidd's [4] principles of modelling, and Ward [5] and Salt [6] also set out a number of advantages of a simple model. However, definitions of level of detail and complexity are not usually provided in the literature and there are no agreed ways of measuring them.

A particularly interesting study in this area is that of Willemain who carried out an experiment to investigate the initial stages of a modelling project [7]. The experiment consisted of providing O.R. experts with a description of an O.R. modelling problem, and asking the expert to speak aloud their thoughts on tackling the problem for a period of an hour, while recording this on tape. Transcripts of the recordings were then analysed by breaking them into "chunks" (from a phrase to a couple of sentences) and categorising each one by a topic in the modelling process. The five topics used were as follows (with the name of the corresponding stage in the simulation process that we would use given in brackets, based on Willemain's descriptions): context (problem structuring), structure (conceptual modelling), realization (data collection and analysis, model coding, model experimentation), assessment (verification and validation), implementation (implementation, including model handover to the client). There were four different problems and four experts tackled all four problems. A further eight experts tackled one problem each giving a total of 24 sessions. The categorisations were analysed in various ways. One of the main results was that even though the sessions only lasted an hour, the experts spent a considerable proportion of the time on all topics other than implementation, with a lot of alternation between the different topics. In particular, structure was often followed by assessment and assessment was often followed by structure. In other words, the experts would tend to think of an aspect of the conceptual model, then evaluate it and then often revise the conceptual model based on this evaluation. Willemain [8] also carried out a survey of the 12 experts which provided revealing insights on their modelling styles and their views on the ideal qualities of modellers, models and clients.

Conceptual modelling is often thought of as a skill that improves with experience. One way for all modellers, but particularly novice modellers, to get better at conceptual modelling is therefore to draw on the experience of experts. Knowledge of what both expert and novice modellers actually do in practice is also an essential foundation for conceptual modelling research. However, apart from Willemain's study [7], there is a lack of empirical studies or data in the literature on how modellers develop conceptual models and on how conceptual modelling relates to the other modelling topics. Section 2 in this paper describes a preliminary study to collect and analyse data on this process.

The overall aim of conceptual modelling is to choose the model that will give the best outcome for the project, which means a combination of various model performance factors such as the confidence in the model, the accuracy of the results, ease of understanding and the resources required to build and use the model (Brooks [9]). Model complexity is a general term that actually has various facets such as model size (the number of model components) and the inter-connectedness of the components, and these are likely to affect model performance in different ways (Brooks [9]). Section 3 in this paper describes an experiment to investigate this relationship between complexity and model performance.

The studies described here are the initial stages of conceptual modelling research and section 4 discusses some of the lessons learnt and possible future work.

2. FOLLOWING MODELLERS

2.1 OBJECTIVE AND RESEARCH APPROACH

The objective was to improve the understanding of the modelling process followed in practice by different modellers, focusing particularly on conceptual modelling. The general approach follows that of Willemain [7] in collecting data on the topics worked on during the modelling process. The studies differ from Willemain's study in four main ways; firstly they are all simulation projects, secondly they are real projects, thirdly data was collected for the whole project rather than just the initial stage, fourthly groups of novices as well as an expert were followed. In fact, moving to real-life projects, looking at novices and looking at groups of modellers were all future experiments suggested by Willemain.

2.2 RESEARCH METHODS

Data was collected for one real project conducted by an expert. The expert has 4 years of modelling experiences in a variety of application fields, including manufacturing, military and health care. The project was carried out part-time by the expert over a period of 10 weeks (Expert project), and involved modelling a call centre to improve the efficiency of staff usage. The modeller was asked to record the total number of hours spent each week on the topics of context, structure, realisation, assessment and implementation used in Willemain's research, as well as the additional topic of experimentation. In Willemain's study the experts did not get as far as actually building and using the model and so experimentation did not occur, although based on his descriptions it would have been included in realization. The Expert modeller was also interviewed each week and asked whether and how the conceptual model had changed during the week and, if the there had been a change, about the process and reasons for changing the model. General questions on, for instance, the main task of the week and whether working on one topic influenced the others were also discussed.

Data was also obtained for six Lancaster University student group projects (Novice project), which lasted for about 12 weeks. Two of the groups were from the simulation module on the MSc Operational Research course and the other four were from the undergraduate simulation course. As part of both courses, the students are required to find a suitable project (typically from around campus) and carry out a complete simulation project. Therefore, although the projects are modelling a real problem there is no external client. The project is a substantial part of the assessment for the courses. The MSc groups had three students whilst the undergraduate groups had five students. Weekly questionnaires were handed out to each group before the project started and each group was asked to record the total hours spent on the different topics every day. In this case a more detailed list of topics was provided than the ones used by Willemain to obtain more detailed data and to reduce the amount of interpretation required by the students. However, these were combined together into the same categories as for the Expert project, with the additional category of report writing, to allow comparison with the Expert project and Willemain's results. In addition, a question concerning conceptual model changes during the week was asked.

2.3 SUMMARY OF THE MAIN RESULTS

The analysis of the data follows some of Willemain's analysis by calculating the relative weights of the different topics, showing a graphical representation of the topics over time and calculating the location of the topics.

(a) Weight breakdown by topics

Figure 2.1 shows the weight of each topic for the Expert project, the Novice projects (average of all 6 projects) and Willemain's experiment (average of all 24 sessions). It is measured in time for the Expert and Novice projects and lines of the transcript for Willemain's experiment.

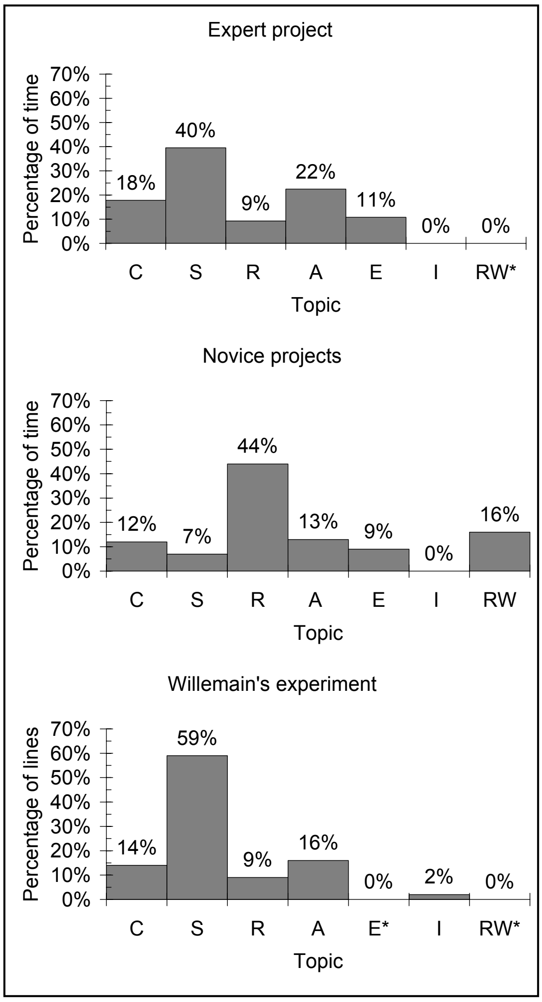


Figure 2.1 Weight of the topics (C = context, S = structure, R = realisation, A = assessment, E = experimentation, I = implementation, RW = report writing, asterisk indicates that topic not recorded separately for that experiment); Willemain's values are from Figure 7 of his paper.

The graphs show a fairly similar weighting of the topics for the Expert project and Willemain's experiment (and the same ordering), but a very different weighting for the Novice projects. One of Willemain's comments on his results was that the initial stages of the projects studied in his experiment were "a little replica of the overall O.R. modeling process", the reason being that the experts considered the role of the model throughout the project as they constructed the conceptual model. The similar weighting of the Expert project to Willemain's experiment is in accordance with Willemain's proposition, although any conclusions must be limited as the Expert data is only based on a single real project.

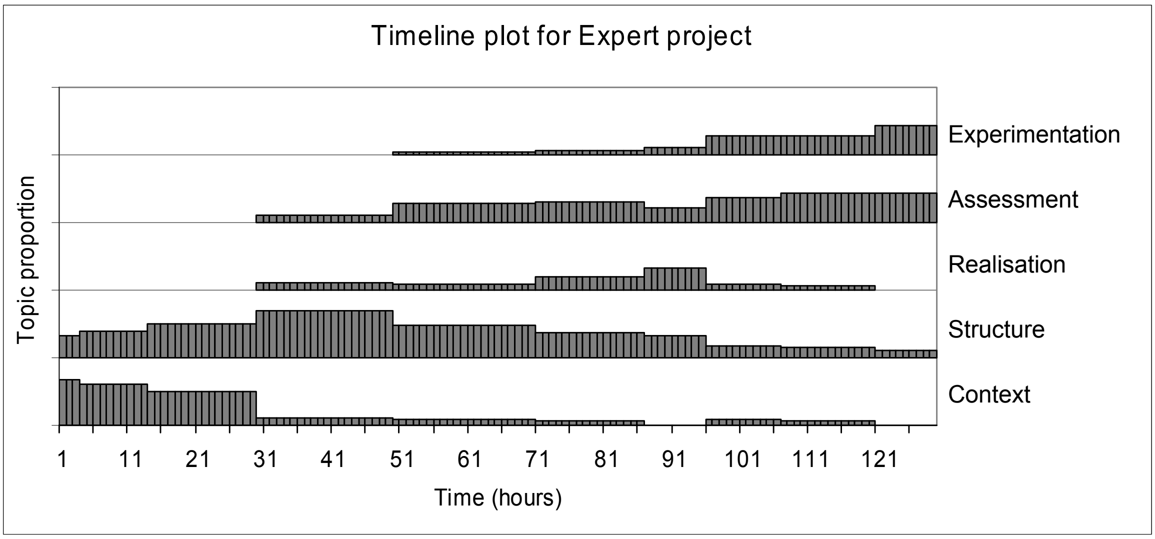
The main differences with the Novice project are the high weighting of realisation and the low weighting of structure. Considering realisation, analysis of the more detailed list of topics provided for the Novice projects showed that some of the novice project groups spent a large amount of time collecting data, whereas this does not apply to the Expert project (the client provided the data) and Willemain's experiment (which is only an artificial exercise involved no data collection). Also, some of the Novice project groups put a considerable amount of effort into model coding which may reflect limited experience of using the software.

The low weighting of structure in the Novice projects indicates a relatively small amount of time spent on constructing the conceptual model and considering alternative models. This could be because the problems tackled were relatively straightforward and so the conceptual model was fairly obvious or it could reflect a tendency of novices to stick with the first model they think of. The greater experience of experts probably means that they consider more alternative models based partly on previous models they have developed.

(b) Timeline plot

The topic timeline plot used by Willemain shows which topic is being worked on at each point during the project. Figure 2.2 shows this type of plot for the Expert project and a typical plot for one of the Novice projects (No. 3). The horizontal axis of the timeline plot represents time in hourly intervals. The vertical axis indicates the topic proportion at each hour.

The Expert project data was obtained on a weekly basis (i.e. the total number of hours spent on each topic in that week) whereas the Novice projects' data were on a daily basis. If more than one topic was worked on in a period then this is shown by the bars not being full height in the plot (a full height bar reaches the horizontal line above on the plot). For example, if a Novice project recorded spending 2 hours on data collection and 3 hours on model coding on a given day then this would be recorded by giving data collection a value of 0.4 and model coding a value of 0.6 over a period of 5 hours. This data collection was less detailed than Willemain's data obtained in a laboratory session, where the protocol recorded what was happening all the time. This is particularly the case for the Expert project where the data was only recorded on a weekly basis (the limitations of this were the reason for changing to a daily basis for the subsequent Novice projects). One consequence is that where more than one topic took place in a recording period then the order and the interaction between the topics is not known. There could have been a lot of switching between the topics or, on the other hand, the topics could have been worked on completely separately one after the other. This prevented a detailed analysis of the switching between topics as carried out by Willemain. Nevertheless, the topic plots still give useful information about positions of the topics throughout the project.



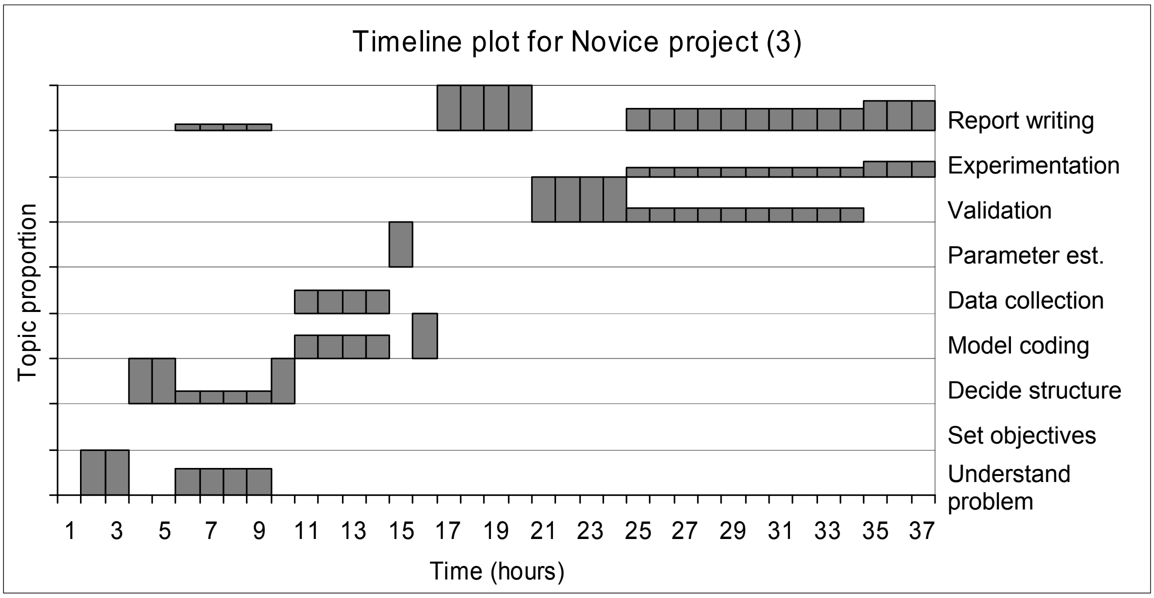


Figure 2.2 Timeline Topic Plots for the Expert Project and a Novice Project

The plots in Figure 2.2 show the topics in the anticipated order with topics higher up on the axis expected to be later. The Novice plot uses the more detailed categories used for this study. Both plots show that generally the topics do tend to appear in the expected order (which was also confirmed by a box plot of position). However, the main difference between the plots is the amount by which the topics overlap. The novice follows a mainly linear process by completing one topic before moving onto the next whereas the expert spends some time on each topic over most of the project duration. In particular, the expert started considering assessment much earlier than the novice modellers and work on assessment frequently caused structure and context revisits. Novice modellers tend to leave model validation and verification till the end of the project and work on assessment rarely renders any reconsideration of appropriateness of model structure. Again this could reflect the novice group tackling a simple problem but is more likely to represent different working practices to the expert.

(c) Change in conceptual models

In both projects, the subjects were asked to note any changes in the conceptual model each week and the reason. For the Expert project, there was one significant conceptual model alteration towards the end of the project. This involved a scope reduction due to the fact that the collected data was not sufficient to support the model built. In one of the Novice groups, one student left in the middle of the project causing a change of project application to make the problem easier to model. The other Novice groups did not adjust the conceptual models after the coding stage started.

3. EXPERIMENT ON RELATIONSHIP BETWEEN MODEL CHARACTERISTICS AND MODEL PERFORMANCE

3.1 OBJECTIVE AND RESEARCH APPROACH

The objective of the experiment was to investigate the relationship between model characteristics and model performance. A better understanding of this relationship would help in the assessment of how alternative conceptual models will perform in the modelling project.

In particular, Brooks [10] suggested splitting model complexity into the more specific characteristics of size (the number of elements used), connectedness (the number of relationships between elements) and calculational complexity (the complexity of calculations determining the relationships). An experiment carried out by Brooks [10] indicated that the difficulty in understanding the model and results is mainly caused by size and connectedness whereas build time is mainly related to calculational complexity. However, that study was a small scale experiment involving 33 M.Sc. students. The approach followed here was to repeat the experimental procedure with different models and to compare the results.

3.2 DESCRIPTION OF THE EXPERIMENT

The experiment used four models built in Witness of a hypothetical planned bank environment. The models were designed to differ in the three aspects of complexity but also had to be quite small so that they could be built by the students within the time constraints of the experiment. Model A is the simplest in terms of measure of complexity and works as the base model. It has a total of 7 different queues and service points and a total of 12 customer routes from one element to another. There are various ways in which the aspects of complexity could be measured. At this stage we will consider size to be the total number of queue and service elements (7) and connectedness to be the ratio of routes to number of elements (1.7). Each of the other models is essentially only more complicated in one aspect of complexity so that the performance difference (if there is any) can be related to a single aspect. Model B only differs from model A by having 2 extra elements (1 queue and 1 service point) and 3 routes to connect them, so that it is larger than model A but with similar connectedness. Model C has the same elements as model A but 5 additional routes so that it differs in connectedness. Model D mainly differs from model A in calculational complexity by having more complex logic within it (service times depend on the route of the customer) although it does have 2 additional routes compared to model A in order to achieve this.

A total of 16 students signed up for the experiment prior to the experiment date, leaving 4 students working on each model. The allocation of the models to students was pre-decided according to their score on a mid-term test so that the effect of variability in student ability on the results could be reduced. This was done by splitting the students into four groups of four students based on ability and allocating the four students in each ability group randomly to a different model.

The experiment consisted of two stages. In the first stage the students were provided with a description of the model and some model output and they were required to answer four questions. The same four questions were asked for each model among which two were quantitative (Q1. calculate the number of customers served, Q3. calculate the average customer waiting time) and two were qualitative (Q2. where to allocate an additional server, Q4. ideas to improve the system). The purpose of asking the participants to answer the questions according to the model description sheet was to test how well they could understand the model. Participants needed to be able to read the statistical results produced by Witness to answer the questions correctly. The students were asked to write the full reasoning for their answers.

The second stage required the participants to build the model in the description sheet in Witness in the shortest possible time. Model build time was recorded at the end.

3.3 DATA ANALYSIS

(a) Analysis of the answers

A marking scheme was first introduced to each question, according to which students' answers were graded. For the first three questions, there were three possible marks of 2, 1 and 0. An integer mark between 0 and 5 was given to question 4. Table 3.1 summarises the average percentage mark for each question for the four different models:

	A	B	C	D	Av.
Q1	50.0	37.5	25.0	75.0	46.9
Q2	75.0	37.5	62.5	75.0	62.5
Q3	62.5	50.0	12.5	50.0	43.8
Q4	62.5	37.5	62.5	25.0	46.9
Av.	62.5	40.6	40.6	56.3	50.0

Table 3.1 Percentage scores for the student's answers.

Overall the participants have a 50% average mark. From the model aspect, model A, the simplest model, produced the highest average percentage mark of 62.5%, followed by model D with 56.3%. Models B and C, more complex in size and connectedness respectively, have the lowest average percentage mark of 40.6%.

Two-way ANOVA and regression analysis were carried out to seek possible relationship between participants' performance (scores) and model type (A, B, C and D) and student type. The regression equations obtained for the scores in the first three questions are listed below. Neither student type nor model type were statistically related (significant at the 5% level) to participants' performance on question 4.

Question 1: Score = 1.96 – 0.43 Student_type

Question 2: Score = 2.36 – 0.41 Student_type – 0.60 Model_B

Question 3: Score = 1.08 - 0.83 Model_C

Student type is a significant determinant for the score for questions 1 and 2, with students in the better ability groups getting higher marks. Size complexity (Model B) also affects the score for question 2, while connectedness (Model C) affects the scores for question 3. Calculational complexity (model D) does not appear to affect the scores. Although the sample size is small the statistical results provide some support for the hypotheses that the model characteristics of size and connectedness affect ease of understanding (as assessed by these questions and for these models), whereas calculational complexity does not. As would be expected, the participants' ability strongly relates to the level of understanding achieved.

(b) Analysis of model building time

Model building time was measured for each model, and the models were also subsequently analysed to identify any errors. In the experiment the students were asked to build the models without carrying out any detailed testing. In general, the performance of the students was quite disappointing with only one participant (A4) managing to build the model without any error.

Various problems were encountered by the students that affected the results. All models had a rule that customers leave the bank if the queue exceeds 15 (denoted here as the Scrap rule since the customers are sent to the "Scrap" destination in the model). Many students had difficulty in modelling this rule and observations during the experiment indicated that those who did not manage to include it spent varying times trying to find a solution. On the other hand some students appeared to simply ignore this rule. Three of the students were unable to connect the elements together (denoted routing problem). None of the model D students modelled the calculational complexity that the service time at a particular counter depends on the type of customer. Instead, participants modelled that each counter server only serves a particular type of customer. It appears that this may be because the model description was unclear and was misunderstood by the students although it could be that the students did not know how to model the required rule.

Table 3.2 shows the results. The number alongside the model denotes the student ability group with 1 being students with best ability and 4 being students with worst ability. Model B4 could not be opened and so was excluded from the results. The time is the number of minutes taken to build the model. The SE column is the number of small minor errors in the model. SR denotes the scrap rule and a 1 in this column means that the scrap rule was included whereas 0 means that it was not. RP denotes the routing problem and here 1 indicates the existence of the problem. LP denotes logic problem and 1 indicates that the model had a significant other logic error. A correct model would therefore have values of 0 1 0 0 in the four columns respectively.

Model	Time (min)	SE	SR	RP	LP
A1	44	2	1	0	0
A2	39	4	0	0	0
A3	51	2	0	1	0
A4	43	0	1	0	0
B1	45	2	1	0	1
B2	40	5	1	0	0
B3	43	1	1	0	1
C1	17	4	0	0	0
C2	16	1	0	0	0
C3	48	4	0	1	0
C4	33	1	0	0	0
D1	41	3	0	0	0
D2	41	1	0	0	0
D3	28	1	0	0	0
D4	52	4	1	1	0

Table 3.2 Model building results (SE = small errors, SR = scrap rule, RP = routing problem, LP = logic problem).

Regression analysis on the data suggested the following model:

Model build time = 39.59 -15.03 Model_C + 15.75 Routing_problem

This equation shows that neither student type nor the number of small errors relate to model build time (based on a 5% significance level). For the three participants who had difficulty in connecting different elements (routing problem), the build time is about 15.75 minutes longer.

The effect on model build time of model C is a reduction by 15.03 minutes. Model C was designed with higher connectedness and so the build time was not expected to be shorter than the build time of the base model A. It is particularly noticeable that C1 and C2 had very low build time. Also none of the model C students included the scrap rule. In fact, the average build time for those who included the scrap rule is approximately 10 minutes longer than whose who didn't model the rule. For those who didn't model the scrap rule, some might have totally forgotten to do so; others might have tried building it, but failed to do so in the end. The very short build time of C1 and C2 could be due to these two students not spending any time considering the scrap rule.

If C1 and C2 are treated as outliers and deleted from the data set then the alternative regression model produced is:

Model build time = 36.83 +11.59 Routing_problem + 5.73 Scrap_rule

Model C is not a significant predictor for this regression model whereas the scrap rule is included. Routing problem still stays as the strongest predictor. This indicates that modelling the scrap rule added an average of 5.73 minutes compared to those that did not model it. The scrap rule involves calculational complexity and, in the absence of the students modelling the complexity of model D, modelling this rule was considered the most difficult part of the modelling.

4. SUMMARY AND CONCLUSIONS

The two studies are preliminary attempts to provide information to improve the understanding of conceptual modelling. One of the main findings from following the Expert and Novice projects is that the pattern for the Expert is similar to Willemain's results for the initial hour of his artificial projects, with a similar relative weighting of topics and the time spent on each topic being spread over most of the overall duration. The Novice projects showed a quite different pattern, with much less time spent on conceptual modelling and quite a linear approach to the project stages. The second study comparing the performance of different models provided some support to the hypothesis that size and connectedness affect model understanding but have little effect on model build time.

However, these findings must be quite tentative as there are limitations to both studies. In the first study the data for the Novice projects was provided by the students themselves. It was hoped to arrange interviews with the students at the end of the projects to discuss the results further but this proved not to be possible. Therefore, the reliability of the data depends on the accuracy of the students in recording the time spent and also on how well they were able to match their tasks against the categories provided. Weekly interviews with the expert provide greater confidence in the Expert project data. More confidence in the data and more detailed data could be achieved if the researcher is present throughout the project and collects the data directly although there are practical difficulties in arranging this. Following real projects also inevitably limits the analysis compared to an artificial laboratory experiment in that all the projects were different. For example, Willemain was able to compare different modellers tackling the same problem and some of the modellers tackling different problems, to try and identify any modeller and problem effects. The second study was limited by the small sample size achieved and also by the difficulties that the students encountered in building the models. The continuation of this research could include repeating these studies but adapting the protocols followed to obtain better data.

Nevertheless, obtaining this sort of information about conceptual modelling and the modelling process is very important in order to improve conceptual modelling and make it more scientific. In the long term, it is hoped that research in this area will improve the success of simulation and O.R. projects, and that the information can help in the training of novice modellers.

REFERENCES

[1] S. Robinson, Simulation: the Practice of Model Development and Use. Wiley, Chichester, 2004.

[2] A.M. Law, "Simulation model's level of Detail Determines Effectiveness", Industrial Engineering, Vol. 23(10), pp. 16-18, 1991.

[3] S. Robinson, "Simulation Projects - Building the Right Conceptual-Model", Industrial Engineering, Vol. 26(9): pp 34-36, 1994.

[4] M. Pidd, Tools for Thinking: Modelling in Management Science, Wiley, Chichester, 1996.

[5] S.C. Ward, "Arguments for constructively simple models", Journal of the Operational Research Society, Vol. 40(2): pp. 141-153, 1989.

[6] J.D. Salt, "Simulation should be easy and fun!" in Winter Simulation Conference. 1993.

[7] T.R. Willemain, "Model Formulation: What Experts Think About and When", Operations Research, Vol. 43(6), 1995.

[8] T.R. Willemain, "Insights on Modeling from a Dozen Experts", Operations Research, Vol. 42(2), 1995.

[9] R.J. Brooks and A.M. Tobias, "Choosing the Best Model: Level of detail, Complexity, and Model Performance. Mathematical and Computer Modelling, Vol. 24(4), pp. 1-14, 1996.

[10] R.J. Brooks, "A Framework for Choosing the Best Model in Mathematical Modelling and Simulation" PhD Thesis, University of Birmingham, 1996.

AUTHOR BIOGRAPHIES

WANG WANG is a PhD student in the Department of Management Science at Lancaster University. She completed her first degree in Management Science (Hons) at Lancaster University in 2003.

ROGER BROOKS is a lecturer in the Management Science department at Lancaster University. He received a PhD and MSc in Operational Research from Birmingham University and a B.A. (Hons) degree in mathematics from Oxford University. He is co-author of a textbook on simulation. Areas of interest include conceptual modelling, Boolean networks and agent-based simulation.