

Evaluating a Rapid Simulation Modelling Process (RSMP) through Controlled Experiments

Rizwan Ahmed*, Tracy Hall*, Paul Wernick*, Stewart Robinson**
University of Hertfordshire UK*, University of Warwick UK**
R.2.Ahmed@herts.ac.uk*, T.Hall@herts.ac.uk*, P.D.Wernick@herts.ac.uk*
Stewart.Robinson@wbs.ac.uk**

Abstract

In recent years there has been an increasing interest in simulation modelling to study, control, and manage software development processes. Despite this, little attention has been paid towards the simulation modelling process itself. We have developed a rapid simulation modelling process (RSMP) based on our study of experienced simulation modellers. This paper reports a set of controlled experiments to evaluate the usability and utility of the RSMP. The results from this show that the RSMP has proved to be a usable and useful approach for developing software process simulation models. The results show that the RSMP is likely to bring discipline to software process simulation modellers and improve the quality of the model they produce.

1. Introduction

In this paper, we present an empirical study to evaluate the effectiveness of a Rapid Simulation Modelling Process (RSMP) for simulation model development of software processes. The RSMP has been developed for novice simulation modellers and is based on a previous empirical study of expert simulation modellers' practices and recommendations [1]. This paper reports the results from controlled experiments designed to evaluate the RSMP.

There is a growing interest in simulation modelling to study, predict, improve, and manage increasingly complex software development processes. Although simulation modelling has been applied very commonly in other scientific and business processes e.g. defence, air traffic control, demand & supply chains, it is relatively new to software engineering. Studies show that simulation modelling has proved to be an effective tool to study software processes and forecast potential change and improvement in those processes [14].

A simulation model is a model that represents the characteristics and features of a system in a dynamic manner [12]. A *software process simulation model* is used to study some particular software activity, such as, development, maintenance or evolution [12].

Despite the increasing interest in software process simulation, less has been said in the literature about the process of simulation model development. The process of developing simulation models is believed to have an effect on the quality of the models produced [5]. In software engineering, simulation modelling has come into research and practice very recently; hence many software process simulation modellers are novices. Our recent survey of software process simulation modellers found that there was a need for a formal process of simulation model development [2]. We found that there is a lack of material to guide the novice software process simulation modeller in how to produce software process simulation model. We have developed a process framework, the RSMP, for novice software process simulation modellers to respond to this need [1]. The experiments reported in this paper are aimed at evaluating the RSMP.

The paper has been organised in 7 sections. Section 2 briefly discusses the RSMP and the research question. Section 3 introduces the research methodology. Section 4 presents the results of the experiments. Section 5 discusses the results and evaluates the usability and utility of the RSMP. Section 6 discusses threats to the validity of the experiments. Finally Section 7 concludes the paper and discusses the future work.

2. Development of the RSMP

We conducted qualitative semi-structured interviews with expert software process simulation modellers and general simulation modellers to explore their practices. The software process simulation modellers simulate software engineering problems, and general simulation modellers simulate business, manufacturing, healthcare and defence problems. The

latter group were involved because simulation modelling is relatively new in software engineering, whereas it is quite an established subject in other disciplines such as operational research and manufacturing. Each group consisted of 10 participants. A study of the two groups allowed comparison and the identification of deficiencies in modelling practices of software process simulation modellers. Our empirical investigation of the practices of simulation modellers resulted in the development of the RSMP, a process framework for novice software process simulation modellers [1].

2.1. The RSMP

The RSMP is an incremental and iterative process for simulation model development. It has three core phases: *Foundation*, *Construction* and *Experimentation* as shown in Figure 1. Each of these core phases has some steps which guide modellers through the modelling process.

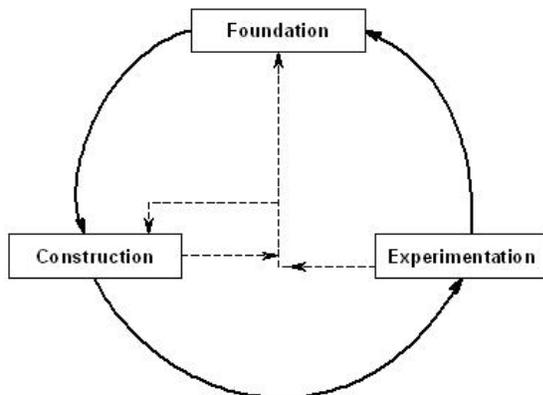


Figure 1: High level description of rapid simulation modelling processes (RSMP)

2.1.1. Foundation. The foundation phase sets up the foundation of the simulation study. It focuses on defining the modelling problem. There are several steps in foundation phase. The simulation modeller will typically communicate the problem with the client.

Problem definition step comes after the problem has been communicated with the client. There are several tasks in the problem definition step. Following a strict order of activities in problem definition is neither necessary nor practical. However, the identification of simulation users should normally be the first step. The rest of the activities are not mutually exclusive, in fact they inform each other. The whole simulation project is driven by simulation goals. A modeller needs to define the goals with the customer clearly and upfront. There can be single or multiple goals of a simulation study. On the basis of simulation goals, modeller will define the questions for which the client needs answers.

System definition includes gaining an understating of the problem by analysing the problem and defining the problem scope. Identifying proper and relevant input variables is very important to simulation. Precise and accurate data increase confidence in the results of a simulation study. Therefore, initial data requirements will be specified to the customer i.e. model inputs. The analysis of outputs from a simulation model answers the questions for which the simulation study has to be carried out. Therefore outputs needed from a simulation model must also be defined. Model output will be defined based on model goals and simulation scope. Technical feasibility will be checked if simulation is appropriate for this kind of problem.

The next step for the modeller would be to decide whether a prototype should be built to gain deeper understanding of the problem under study. Talking through initial prototype with the client may further inform the problem definition. The next step would be tool/technology selection depending on the type of problem being simulated.

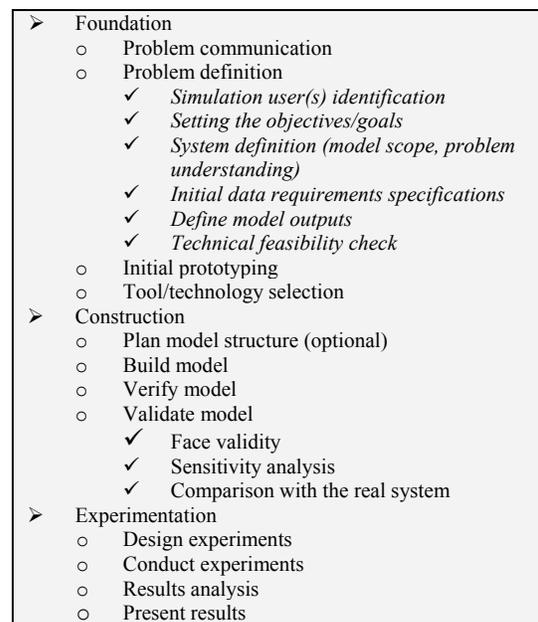


Figure 2: The RSMP

2.1.2. Construction. In construction phase, actual physical simulation model would be generated on computer using some simulation tool or programming language.

First of all the model structure should be planned. Different modellers may have different ways to structure the model. RSMP does not insist to produce a detailed design of the simulation model. Structure of the model may be planned on a piece of paper or using a CASE (the computer aided software engineering) tool, depending on the nature of the problem.

Then start building the model. This may be one sub-model or module or it may build on the initial prototype produced in Foundation phase.

Then the computer model will be verified. Verification is a micro level check on the model behaviour i.e. the modeller will check that correct logic has been employed. It is similar to software code debugging. If any problem is found in the model, the modeller may have to go back to the previous step and make changes in the computer model

The next step is validating the model. Validation checks whether you have built the right model. The purpose of validation is to ensure that model behaves correctly overall. There are two dimensions to model's validity; one from the modeller's perspective and the other from the client's perspective (often called credibility). In any model validation the client must be heavily involved. There are various validation techniques e.g. face validity, sensitivity analysis, and comparison with real system. If the model is invalid, the modeller may have to revisit foundation phase and then change computer model as necessary.

2.1.3. Experimentation. Experimentation phase allows a modeller to provide the client with the real value he/she wants from a simulation study.

The experimental design is driven by the goals, questions and scope of the problem in the simulation study. Different simulation run scenarios are discussed with client, under which simulation has to be performed. Experiments are designed on the basis of these scenarios. The rigour of experimentation depends on the scope and goals of the problem.

The results obtained from the experiments runs are analysed. At this point some changes may have to be made in the simulation model depending on the kind of results obtained. Under certain situations, client may run experiment and analysed results herself. The client may ask to change or add features the model on the basis of that analysis. Then modeller may have to revisit Foundation and/or construction phases

Once the results have been analysed they are put in a presentable format to facilitate the client with decision making. Results may be put in graphs or tabular formats. Any statistical analysis is discussed and conclusion provided.

2.2. Research question

The aim of the controlled experiments reported in this paper is to answer the research question:

RQ: *Will an empirically developed simulation modelling process help novice software process simulation modellers to improve their simulation modelling?*

These experiments aim to answer the research question by evaluating the:

- Usability of the RSMP; how easy is it to follow the RSMP
- Utility of the RSMP; how useful the RSMP is

3. Research methodology

We decided to conduct controlled laboratory experiments to gain greater control. This was mainly because the performance of the participants had to be measured under similar conditions and compared so as to study the difference in their performances using the RSMP and not using the RSMP.

3.1. The Sample

Participants were selected from the MSc. Software Engineering module, Models and Measures in Software Engineering (MMSE) at the School of Computer Science, University of Hertfordshire.

The MMSE class, consisting of 45 students, was given three lectures, 6 hours of class contact, in system dynamics modelling as part of this module. After these three lectures a class test, which was part of the course, was held. Based on the scores obtained in the test, the top 10 students were invited to participate in the experiments. None of the invited participants had any prior experience of simulation modelling. Each participant was offered a 10 pound incentive to take part in the experiments. All 10 students agreed to participate in the experiment; however, one participant did not turn up on the day of the experiment.

The participants were divided into a control group (the Non-RSMP group) and a treatment group (the RSMP group). To ensure that both groups have a balanced skill level, Participants were assigned to the groups on the basis of the marks they obtained in the class test. They were pair matched.

Five students participated in the Non-RSMP experiments and four students participated in the RSMP experiments. However, one participant in the Non-RSMP group could not work properly because of some technical problems. This leaves four participants in each group for the analysis. Participants of the Non-RSMP group will be referred to as A1, A2, A3, A4 and participants of the RSMP group will be referred to as B1, B2, B3, and B4 throughout this paper.

3.2. Experimental setting and protocol

To emulate the real environment of modeller and client, one person from the research team acted as a client. Participants had free access to the client for as

much time as required to emulate a modeller-client relationship.

Each participant of the RSMP group was tutored in the RSMP individually. Care was taken to ensure that tutoring the RSMP should not increase their skills in system dynamics modelling so as to keep the skill level between the groups balanced. For this purpose, the discussion in the RSMP tutorial was focused on general process steps without referring to system dynamics modelling. Each participant was then given a questionnaire to evaluate their understanding of the RSMP.

The duration of each experiment was 2 hours. Given that students had other commitments and pressures, this duration was short enough but realistic enough to allow the use of RSMP.

A case study adapted from Misić et al. [13] was given to participants to model. In the first session of experiments, the Non-RSMP group used their own personal approach to modelling. This was based on general material covered in class lectures. In the second session of experiments, the RSMP group used the RSMP to develop the simulation model for the case study.

Each participant was provided with a computer installed with the Vensim tool (www.vensim.com) for system dynamics modelling. Each participant was provided with the case study two days prior to the experiments. Each participant's activities on the computer were recorded using screen capture software, Screen Movie Studio (www.mandsoft.com), to measure the effort spent on different tasks and capture the pattern of their activities.

At the end of the experiment, each participant was given a questionnaire to evaluate the experiments from the participant's perspective. Moreover, RSMP participants were given a questionnaire to evaluate their experience of using the RSMP.

The models produced by the participants have been assessed using evaluation criteria, which were established from simulation literature recommendation earlier in the research program [1, 13]. Devising the evaluation criteria prior to developing a simulation modelling process and conducting the experiments minimises bias.

The evaluation criteria assess the models for their syntactic quality, semantic quality, design, quality of documentation and maintainability. The syntactic quality of each model was assessed by a member of the research team. Semantic quality, design and the quality of documentation were evaluated by the client. Maintainability was assessed by the participants by swapping the models across the two groups.

A pilot experiment was conducted to validate the approach with one pilot participant prior to the main set

of experiments. The experience of pilot gave general support to the methodology. To increase confidence in the questionnaires designed for these experiments, they were also validated in the pilot experiment. The pilot evaluation suggested minor wording changes in the questionnaires.

4. Results

In this section, we present the results of these experiments based on the artefacts produced by the participants and our research observations. Moreover, participants' patterns of activities will also be discussed and compared.

4.1. Syntactic correctness

Two things were evaluated in terms of syntactic correctness. First, the correctness of the model diagrams. This was automatically checked by the modelling tool. Second, the correctness of the mathematical equations employed in the model. Most of the participants in both groups could not get the equations right. Participants were asked in the experiment evaluation questionnaire to comment on how happy they were with the model they produced. It showed that none of the participants were completely happy with the simulation model they produced. Discussion with the participants suggested that this is because of their lack of experience with system dynamics modelling and the modelling tool. None of the participants was an expert simulation modeller. They were not confident with the modelling tool because they had limited experience of using Vensim. As none of them produced a complete working model, the evaluation of syntactic correctness of the models is not possible.

4.2. Semantic quality and design

The client was provided with a questionnaire to evaluate the quality of the simulation models produced. The results are shown in Table 1.

The models' semantic quality was evaluated in terms of 'face validity and scope coverage'. The design of the model was evaluated in terms of 'modularity, interoperability and clarity'. Table 1 shows that according to the client evaluation, only one participant, A1, from the Non-RSMP group and two participants, B1, B2, from the RSMP group produced valid models at face value. A1 from the Non-RSMP group and B1 and B2 the RSMP group produced models which cover the scope of the given problem. The client was satisfied with the modularity of the models produced by one Non-RSMP participants, A1, and two RSMP

participants, B1 and B2. A1 produced a model which the client considered interoperable compared to none from the RSMP participants. A1, B1, B2 and B4 produced models which the client considered clear enough in their layout.

Verification and validation of simulation models was also part of the evaluation of semantic quality. This was to be evaluated by both the participant and the client. However, as none of the participants could produce a complete working model, verification and validation of the simulation models could not be completed.

4.3. Quality of the documentation

The client was provided with a questionnaire to evaluate the quality of documentation produced to support the models. Results are summarised in Table 2.

All the RSMP participants provided some amount of documentation. The client was not satisfied with the documentation provided by any of the Non-RSMP participants because they provided little or no documentation. According to the client, the best documentation was produced by the RSMP participants, B3 and B4, as shown in Table 2.

4.4. Maintainability

The maintainability of the simulation models produced by the two groups was assessed. The simulation models and their accompanying documentation were swapped across groups and participants were provided with a questionnaire. Results have been summarised in Table 3.

The Non-RSMP group gave the RSMP group a very good score for model documentation compared to what they obtained from the RSMP group. Analysing participants' responses to this questionnaire suggests that the RSMP group scored better than the Non-RSMP group. This implies that simulation models produced by the RSMP group are more maintainable than those of the Non-RSMP group. We can relate this to the fact that the RSMP group spent a lot of time (on average 17% of their time) on model documentation.

Also the RSMP group obtained much better scores for model structure than that of the Non-RSMP group. However, the Non-RSMP group scored slightly better for meaningfulness of variable names and comments than the RSMP group. This may be related to the fact that the Non-RSMP group spent more time using the modelling tool (on average 61% of their time) than the

RSMP group; therefore, they had more time to name and review their variables and comments. An analysis of their recorded activities on computer provides evidence that they renamed and reviewed their variable names and comments several times.

4.5. Participants' effort distribution and pattern of activities

Participants' activities on the computer were recorded using screen capture software and the duration and number of client contacts for each participant was recorded. This allowed the effort spent by each participant on different tasks and their pattern of activities to be calculated. A comparison of effort distribution and activities patterns across the two groups allowed the correlation of various factors. These factors are now discussed.

4.5.1 Client contact. Table 4 summarises the client contact time of each participant. It shows that the duration and the number of client contacts of the RSMP participants are much higher compared to the Non-RSMP participants. The Non-RSMP participants spent most of their time working with the modelling tool, whereas the RSMP participants spent a lot more time with the client.

The mean time spent by the Non-RSMP group on client contact is 11.25 minutes, 9 percent of total time on average, whereas that of the RSMP group is 22 minutes, 20 percent of total time on average. A *t*-test to compare the means of the two groups' client contact time confirms that this is a statistically significant result.

4.5.2. Tool use. Table 5 summarises tool use of the participants. The analysis of screen capture data for tool use shows that the Non-RSMP participants preferred to start working with the tool as soon as they saw the client. Each Non-RSMP participant started working with the tool within 12 minutes of first seeing the client. While the earliest an RSMP participant started working on the tool was the 33rd minute after seeing the client.

The average time spent by the Non-RSMP group on the tool is 74 minutes, 61 percent of total time on average, whereas that of the RSMP group is 37 minutes, 31 percent of total time on average. Again a *t*-test to compare the two groups' tool usage time confirms a statistically significant difference.

Table Notes

- *Responses on 5 point likert scales; Strongly Disagree 1 2 3 4 5 Strongly Agree
- M = Median of the corresponding values
- Time is in minutes in all tables

Table 1: Client evaluation of model's semantic quality and design

	A1	A2	A3	A4	M	B1	B2	B3	B4	M
1. The model is valid on face value	5	3	1	3	3	4	4	2	3	3.5
2. Model feasibly covers the scope	5	2	2	3	2.5	5	4	3	2	3.5
3. The model is feasibly modular	4	1	2	2	2	4	4	1	2	3
4. It would be easy to couple this model with another model	4	2	3	3	3	3	3	1	3	3
5. Model layout is clear enough	5	2	2	3	2.5	4	4	2	4	4

Table 2: Client evaluation of the documentation

	A1	A2	A3	A4	M	B1	B2	B3	B4	M
1. Simulation objectives have been defined well	1	1	1	1	1	3	1	5	5	4
2. Simulation questions have been defined well	1	1	1	1	1	3	1	5	5	4
3. Model scope has been defined well	1	1	1	1	1	3	1	4	5	3.5
4. Model inputs have been defined well	1	1	1	1	1	2	1	4	4	3
5. Model outputs have been defined well	1	1	1	1	1	2	1	4	4	3
6. Overall documentation is good	1	1	1	1	1	2	1	5	5	3.5

Table 3: Peer Evaluation

	A1	A2	A3	A4	M	B1	B2	B3	B4	M
1. Model documentation is well structured	1	1	1	3	1	2	1	4	4	3
2. Model documentation is helpful to understand the model	1	1	1	3	1	3	1	4	4	3.5
3. Model structure is easy to understand	2	3	1	2	2	4	3	2	4	3.5
4. Model has meaningful variable names	4	2	2	4	3	4	2	2	3	2.5
5. Comments in the model help me to understand the model	3	1	1	3	2	1	1	1	3	1

Table 4: Client contact

	A1	A2	A3	A4	B1	B2	B3	B4
Number of times seen the client	2	2	2	1	6	3	5	2
Client contact Time	17	13	12	3	31	15	27	15
Mean Time spent on client contact	11.25				22			
Statistical significance value	0.039 < 0.05 Difference is statistically significant							

Table 5: Tool usage

	A1	A2	A3	A4	B1	B2	B3	B4
Tool usage Time	87	59	72	79	34	50	N/A*	25
Average Time spent on tool usage	74				37			
Statistical significance value	0.005 < 0.05 Difference is statistically significant							
N/A* = B3 did not produce model on tool, hence B3 is not included in calculating avg. time of tool usage								

4.5.3. Documentation. The RSMP participants spent a substantial amount of their time on documentation but Non-RSMP participants spent a negligible time on documentation as shown in Table 6. The RSMP group

spent 20.5 minutes, 17 percent of total time on average on documentation compared to 1.25 minutes, 1 percent of total time on average for the Non-RSMP group. A *t*-test to compare the means of the two groups' documentation time confirms a statistically significant difference. This difference of effort on documentation suggests that Non-RSMP participants either did not consider documentation important or they were so bogged down in the technical work that documentation got ignored.

Table 6: Documentation time

	A1	A2	A3	A4	B1	B2	B3	B4
Total time	0	0	0	5	26	6	30	50
Average Time spent on tool usage	1.25				20.5			
Statistical significance value	0.013 < 0.05 Difference is statistically significant							

4.5.4. Other activities. Other activities include time spent on rough work and thinking. Table 7 shows the effort distribution of participants on other activities. There is no way to separate the amount of time spent by the participants on rough work and thinking, therefore, these are combined. Analysis suggests that a big difference in miscellaneous activities time cannot be observed between two groups. On average the time spent by the Non-RSMP group on miscellaneous is

lower (24.5 minutes) than that of the RSMP group (31.67 minutes).

The rough work produced by the participants included any preparation done by the participants before coming to the experiments and any rough work done during the experiments. These rough documents were retained after the experiment. Table 8 summarises different themes identified in the rough work produced by the participants in both groups. Consistency in the activities is observable amongst the RSMP participants. The rough work by A3 shows that A3 naturally performed the activities suggested by the RSMP even though A3 was a Non-RSMP participant. However, looking at the rough work, it is not possible to identify the sequence of these activities performed by the participants. An interesting finding is that those participants who have tried formulating equations on paper have tried employing equations in the actual physical model.

Table 7: Other activities

	<i>A1</i>	<i>A2</i>	<i>A3</i>	<i>A4</i>	B1	B2	B3	B4
Paperwork and thinking Time	16	34	15	33	29	49	63	17
Average Time spent paper work and thinking	24.5				31.67			
Statistical significance value	0.12 > 0.05 Difference is Not statistically significant							

Table 8: Rough work

	<i>A1</i>	<i>A2</i>	<i>A3</i>	<i>A4</i>	B1	B2	B3	B4
1. Simulation questions	<i>Y</i>	<i>Y</i>	<i>Y</i>	<i>N</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>
2. Inputs identification	<i>Y</i>	<i>N</i>	<i>Y</i>	<i>N</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>
3. Influence diagram	<i>N</i>	<i>Y</i>	<i>Y</i>	<i>N</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>
4. Model doodling	<i>Y</i>	<i>N</i>	<i>Y</i>	<i>N</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>
5. Equations formulation	<i>Y</i>	<i>N</i>	<i>Y</i>	<i>N</i>	<i>N</i>	<i>N</i>	<i>Y</i>	<i>N</i>
6. Outputs identification	<i>Y</i>	<i>N</i>	<i>N</i>	<i>N</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>
Y = yes, N = No								

5. Discussion

This section discusses the results of the experiment in terms of the usability and utility of the RSMP and relates them to the literature.

5.1. Usability of the RSMP

This section discusses how well the experiment participants were able to follow the recommended guidelines of the three key areas of the RSMP. The three key practice areas are:

- Client contact
- RSMP core phases (foundation, construction and experimentation)
- Documentation

The RSMP puts emphasis on heavy client contact. The results show that all the RSMP users spent a fair amount of time, 20 percent of the total duration compared to that of 9 percent of the Non-RSMP participants, with the client. This shows that the participants were able to follow the RSMP guidelines of having heavy client contact. This confirms the literature recommendation that having a disciplined simulation modelling process is likely to increase communication with the client [7, 15].

RSMP has three core phases (foundation, construction and experimentation). The RSMP participants spent about 47 percent of their time on average at the foundation phase of the RSMP. The documentation provided by the RSMP group shows that they adhered to the RSMP foundation phase guidelines. Also the consistency in the paperwork activities of the RSMP participants in Table 8 provides evidence that they tried adhering to the RSMP foundation phase guidelines. All of them established the questions to be answered by the simulation study, identified inputs and outputs, and derived an influence diagram showing the relationships between different factors in the problem. Moreover, most of the RSMP participants defined the system and its scope under study in their documentation. The RSMP participants spent 31 percent of their time on average on the RSMP construction phase. All the RSMP participants planned the model structure on paper prior to building the physical model, as instructed by the RSMP. However, they did not do any experimentation with the model. This is because they could not produce a complete working model due to the lack of experience with system dynamics and the modelling tool. This should not be seen as a weakness in the RSMP, rather a weakness in the selection of participants. Analysis of the data collected suggests that RSMP participants were able to follow the first two phases.

Providing the model documentation is another RSMP key practice. RSMP participants were encouraged to produce the documentation as they model. They spent 17 percent of their time on average on documentation. Although the client was satisfied with the documentation produced by only two of the

RSMP participants, each participant in the RSMP group did spend some time on documentation.

This discussion of the findings suggests that the RSMP is a usable approach. Participants found it easy to understand and adhere to the RSMP guidelines. . Therefore it can be concluded that the RSMP was an easy to use process for the participants.

5.2. Utility of the RSMP

The utility of the RSMP has been evaluated on the evaluation criteria:

- Semantic quality
- Model design
- Quality of documentation
- Maintainability

Because none of the participants could produce a completely working model, it was not possible to apply all the evaluation criteria to the model produced by the participants.

The evaluation of each model's semantic quality included its credibility, scope, verification and validation. Unfortunately verification and validation of the simulation models was not possible because the models were not fully functional. However, their credibility was evaluated by the client judging them for face validity. The client evaluation suggests that overall the RSMP participants produced more credible models than the Non-RSMP participants. This is inline with Gass and Joel's [8] suggestion that discipline in modelling process is likely to produce credible models. Moreover, the models produced by the RSMP participants cover the scope better than those of the Non-RSMP participants.

The client evaluation shows that overall the models produced by the RSMP participants were more modular and clear in their layout as compared to those of the Non-RSMP participants. This may be related to the fact that the RSMP group spent more time interacting with the client and doing paper work (on average 47% of their time). This allowed them to think methodically about the model structure, therefore, they were able to produce a better model structure than the Non-RSMP group. This confirms the literature finding that planning the model structure in advance is likely to produce better model design [16]. However, evaluation of the interoperability of the models produced by participants does not indicate any significant difference between the two groups.

Literature suggests that heavy client contact promises improved model quality [7, 15]. However, looking at the evaluation results, it would be hard to relate client contact with the quality of the produced simulation models. According to the client evaluation, the best model was produced by a Non-RSMP

participant, A1, whose client contact time was only 17 minutes, which is much lower than that of B1, who spent 31 minutes with the client and produced the best model in the RSMP group. Perhaps, this result is related to the lack of sophistication in the modelling problem. However, A1 was also the top scorer in the system dynamics class test. Whereas, both the RSMP participants, B1 and B2, with whose models the client was satisfied scored lower in the class test. This suggests that the RSMP training equipped them with extra potential to produce a model that could satisfy the client.

Elmaghraby [6] says that simulation modellers tend to spend too much time on building the model rather than developing an understanding of the problem to be modelled. It is interesting that the Non-RSMP participants spent most of their time, 61 percent of the total on average, on the tool but only one Non-RSMP participant could produce a model that could satisfy the client. In contrast to that, the RSMP participants spent much less time, only 30 percent of the total time on average, on the tool but two of the RSMP participants satisfied the client evaluation. Therefore, this suggests that spending lots of time on the tool did not add significantly to the quality of the simulation models produced in these experiments.

Documentation is a concern that Gass [7, 10] believes can be solved by strengthening the simulation modelling process. The results about documentation evaluation show that the more time spent on documentation, the better the documentation produced. Interestingly in the RSMP group, B3 and B4, who provided excellent documentation, did not produce good enough models to satisfy the client. One explanation of this can be they were too process focused; hence they could not produce a good model. Whereas, the two RSMP participants, B1 and B2, produced satisfactory models, could not produce good enough documentation to satisfy the client. This suggests that the participants may not have balanced their focus between product and process. The client evaluation for the documentation produced by the two groups suggests that the RSMP participants produced much better documentation than the Non-RSMP participants. This is primarily due to the difference of approaches adopted by two groups for developing the simulation models.

Maintainability of the models was evaluated by swapping the models and their documentation across two groups. Peer evaluation of the models' maintainability, as shown in Table 3, suggests that the model produced by the RSMP participants were more maintainable than those of the Non-RSMP participants. It is important to note that documentation is believed to be important to understand and hence maintain

simulation models [3, 9]. Therefore, this can be related to the fact that RSMP participants provided better documentation hence their models were considered more maintainable by their peers. However, the abilities of the peer to evaluate the model may be questionable, as the participants in both groups did not have significant simulation modelling experience. Therefore, one may argue that participants evaluated their peer's model's maintainability intuitively rather than basing this on a deep understanding of the simulation modelling technique.

One weakness of this study, which limits justifying the utility of the RSMP, is that no validation of the models could be performed.

Overall our results suggest that the RSMP participants were able to produce better models than the Non-RSMP participants. The client was satisfied with the models produced by two of the RSMP participants compared to one Non-RSMP participant. Moreover, the client was satisfied with the documentation produced by two of the RSMP participants compared to none of the Non-RSMP participants. From the maintainability perspective, the RSMP participants scored better than the Non-RSMP participants.

5.3. Evaluation of the RSMP: Conclusion

The two objectives, evaluating the usability and the utility of the RSMP, to answer the research question have been partially satisfied through these experiments. The research question aims to investigate the improvement in simulation modelling practice of novice software process simulation modellers using the RSMP. The results show that the RSMP has proved to be usable and useful for the participants in some aspects, however, the RSMP's utility remains questionable in terms of simulation model validity. Another empirical study, involving participants with more training, will now be conducted to justify the utility of the RSMP for simulation model validity.

Although this study does not indicate what effect the RSMP may have on the validity of the models, nevertheless, the findings from this study suggest that the RSMP may bring benefits in other important areas such as, problem formulation, model design, documentation, and maintainability. This can be augmented by participants' perceptions about the RSMP. In the RSMP pre evaluation questionnaire, RSMP participants were not sure if following the RSMP would improve the quality of the model they produce. However, in the post evaluation they found it was useful at improving the quality of their models. It is interesting that most of the RSMP participants perceived that using the RSMP is likely to bring discipline in their simulation modelling practice and

they had a similar perception in the RSMP post evaluation. Similarly they scored the same in both evaluations for the helpfulness of documentation guidelines; they perceived and found RSMP documentation guidelines useful. Therefore, the results from these experiments confirm the usability and utility of the RSMP in certain areas for these participants.

6. Threats to the validity of experiments

There are several potential threats identified to the validity of these experiments.

The representativeness of the sample is limiting. We could have chosen participants randomly from the population of students in the class, making the sample representative of the student population in statistical terms. Choosing the best students was aimed at generating a sample as close to novice simulation modellers as possible in terms of skills. Therefore the sample is representative of the best skills in the population of students.

A major weakness in these experiments was that the participants were not able to produce working models which limited the evaluation of the RSMP. A follow up questionnaire results revealed that lack of expertise with system dynamics and the modelling tool was the main cause of failure to produce a working model. This suggests that the participants should have been provided with more training in system dynamics and the modelling tool. To eliminate this threat, the participants will be given more sophisticated training in simulation modelling and the modelling tool in a future study.

The difficulty of the modelling problem could have also affected the performance of the participants. However, none of the participants indicated that it was the difficulty of the problem that hindered him/her producing working models.

There was a two week lag between the Non-RSMP and the RSMP experiment sessions. The RSMP participants could access the problem statement from the Non-RSMP participants and work on the problem during those two weeks. To mitigate this risk, Non-RSMP participants were asked not to show the problem statement to the RSMP participants. Moreover, the RSMP participants could self study system dynamics and improve their skills over the Non-RSMP participants. This was an unavoidable risk, as we had no control over their activities. However, considering participants' other commitments this seems unlikely.

There are no standard evaluation criteria in practice to evaluate a simulation model. The evaluation criteria were drawn through an analysis of the simulation modelling literature. The validity of criteria used for the

evaluation of simulation models can be another potential threat.

The short duration of experiments may be another potential threat to the validity of the conclusions drawn from these experiments. A field study of the RSMP with practitioners, where a simulation study may last for weeks or months, would add to the validity of the conclusions from these experiments.

RSMP is a generalised process applicable to both discrete event and continuous simulation. Ideally it should have been assessed for a discrete event simulation problem as well; however, resources and time issues constrained us to test it only with continuous simulation. It would also be assessed with a discrete event simulation problem in future to further generalise the validity of the RSMP.

7. Conclusion and future work

This study of controlled experiments was aimed at investigating the usability and utility of the RSMP in a specific context. This paper concludes that the RSMP has proved to be 'partially' usable and useful for the participants of these experiments. The RSMP participants have performed better than the Non-RSMP participants in these experiments. The purposeful sampling and a small sample size limits the generalisability of the results to a different sample. However, the participants in these experiments had the highest system dynamics skills of their class. This was the closest sample of participants we could identify to novice software process simulation modellers. Therefore, similar results may be anticipated repeating these experiments with novice software process simulation modellers.

While it would not be possible to evaluate the RSMP through a single study, we have designed a multi-phased laboratory study to evaluate the RSMP. This would allow us refine the RSMP and the subsequent phases of the study based on the experience from the previous study. The analysis of these experiments provided us with the opportunity to refine the RSMP. The next stage of this multi-phased evaluation plan aims to evaluate the RSMP through expert judgements. The RSMP will be then further refined and improved based on expert judgements and recommendations. The third phase of the study will be a replication of these experiments with better skilled participants to evaluate the improved version of the RSMP.

Reference:

- [1]. R. Ahmed, T. Hall, P.D. Wernick, and S. Robinson, "Developing a rapid simulation modelling process (RSMP)", TR427, School of Computer Science, University of Hertfordshire. 2005
- [2]. R. Ahmed, T. Hall, and P.D. Wernick, "Practices of ProSim03 modellers: A survey", *Proceedings of ProSim04 workshop*, Edinburgh, UK, May 2004. pp. 67-76
- [3]. O. Balci, "Requirements for Model Development Environments", *Computers and Operations Research*, vol. 13, 1986, pp. 53-67.
- [4]. Bernard, H.R. "*Social research methods qualitative and quantitative approaches*", Sage Publications, Inc. Thousand Oaks, California. (2000).
- [5]. D.M. Eriksson, "A framework for the constitution of modeling processes: A proposition", *European J. of Operational Research*, Elsevier Ltd, (2003), Vol. 145, pp.202-215
- [6]. S. E. Elmaghraby, "The role of modelling in IE design". *Industrial Engineering* Vol. 19, Issue. 6, (1968), 292-305
- [7]. S.I. Gass, P.F. Roth, and A.J. Lemoine, "Some consideration for improving federal modelling", *Proceedings of 1978 Winter Simulation Conference*.
- [8]. S.I. Gass, and L.S. Joel, "Concepts of model confidence." *Computers & Operations Research*. Vol. 8, No. 4, (1981), pp. 341-346
- [9]. S.I. Gass, "Decision aiding models: Validation, assessment, and related issues for policy analysis." *Operations Research*, Vol. 31, No. 4, (1983), pp. 603-631
- [10]. S.I. Gass, "Managing the Modelling process: A personal reflection", *European Journal of Operational Research*. Vol. 31, (1987), pp. 1-8
- [11]. R.J. Brooks, and A.M. Tobias, "Choosing the best model: Level of detail, complexity, and model performance", *Mathl. Comput. Modeling*, Elsevier Science Ltd., 1996, Vol. 24, No. 4, pp. 1-14
- [12]. M. Kellner, R. Madachy and D. Raffo, "Software process simulation modeling: Why? What? How?", *Journal of Systems and Software*, 1999, pp.91-105
- [13]. V.B. Mistic, H. Gevaert, and M. Rennie, "Extreme dynamics: Towards a system dynamics model of the extreme programming software development process", *proceedings of ProSim Workshop*, May 2004, Edinburgh, UK.
- [14]. M. Kellner, and D. Raffo, "Empirical analysis in software process simulation modelling", *Journal of Systems and Software*, Vol. 53, No. 1, 2000, pp. 31-41
- [15]. T.R. Willemain, "Insights on modelling from a dozen experts", *Operations Research*, Vol. 42, No. 2, (1994), pp213-222
- [16]. D. Pfahl, and G. Ruhe, "IMMoS: A methodology for integrated measurement, modelling and simulation", *Software Process Improvement and Practice*, John Wiley & Sons, Ltd. Vol. 7, (2002), pp.189-2