Kent Academic Repository
Full text document (pdf)

Citation for published version

Monks, T and Robinson, S and Kotiadis, Kathy (2014) Learning from discrete-event simulation: Exploring the high involvement hypothesis. European Journal of Operational Research, 235 (1). pp. 195-205. ISSN 0377-2217.

DOI

Link to record in KAR

https://kar.kent.ac.uk/51025/

Document Version

Author's Accepted Manuscript

Learning from Discrete-Event Simulation: Exploring the High Involvement Hypothesis

Thomas Monks[a], Stewart Robinson[b] and Kathy Kotiadis[c]

[a]University of Exeter Medical School

University of Exeter

Exeter

EX1 2LU

UK

Corresponding Author

Email: thomas.monks@pcmd.ac.uk

Fax: 01392 421009

Phone: 01392 726086

[b]School of Business and Economics,

Loughborough University,

Loughborough,

Leicestershire,

LE11 3TU, UK

[c]Warwick Business School

University of Warwick

Coventry

CV4 7AL

UK

ABSTRACT

Discussion of learning from discrete-event simulation often takes the form of a hypothesis stating that involving clients in model building provides much of the learning necessary to aid their decisions. Whilst practitioners of simulation may intuitively agree with this hypothesis they are simultaneously motivated to reduce the model building effort through model reuse. As simulation projects are typically limited by time, model reuse offers an alternative learning route for clients as the time saved can be used to conduct more experimentation. We detail a laboratory experiment to test the high involvement hypothesis empirically, identify mechanisms that explain how involvement in model building or model reuse affect learning and explore the factors that inhibit learning from models. Measurement of learning focuses on the management of resource utilisation in a case study of a hospital emergency department and through the choice of scenarios during experimentation. Participants who reused a model benefitted from the increased experimentation time available when learning about resource utilisation. However, participants who were involved in model building simulated a greater variety of scenarios including more validation type scenarios early on. These results suggest that there may be a learning trade-off between model reuse and model building when simulation projects have a fixed budget of time. Further work evaluating client learning in practice should track the origin and choice of variables used in experimentation; studies should also record the methods modellers find most effective in communicating the impact of resource utilisation on queuing.

1 Introduction

It is often assumed that clients of simulation experience significant learning through involvement in model building (Robinson, 2004; Rouwette, Korzilius, Vennix, & Jacobs, 2011). Although this *high involvement hypothesis* is plausible, it is difficult to measure client learning in practice and little empirical evidence exists to validate the theory. Exploration of the role involvement in model building plays in client learning is important. Particularly as the effort needed to build a discrete-event simulation (DES) model may affect study feasibility or limit scope due to a fixed budget of time (Cochran, Mackulack, & Savory, 1995; Law, 2007; Pidd, 2004; Robinson, Nance, Paul, Pidd, & Taylor, 2004). In fact, given a fixed budget of time, a modeller may choose not to involve clients in model building, but instead reuse an existing or generic simulation model (Bowers, Ghattas, & Mould, 2012; Fletcher & Worthington, 2009; Robinson, et al., 2004). For example, the time saved by reusing a model of a whole hospital could instead be used for experimentation (Günal & Pidd, 2011) or, where no time is available for model building, pre-built models could be used to rapidly educate clients in approaches to improvement, for instance in lean (Robinson, Radnor, Burgess, & Worthington, 2012). Given that reuse is occurring in practice, it is important to understand how client learning is influenced by the reduced involvement in model building and the increased opportunity for experimentation offered by model reuse.

To address this issue, this paper details a laboratory experiment where learning is measured using Argyris and Schön's (1996) theory of action and learning loop framework. We seek to investigate if the effect can be demonstrated empirically; to understand the mechanisms that aid client learning from involvement in building and reuse; and to explore the factors that inhibit learning from simulation. Comparisons are made between the learning novice simulation clients (undergraduate business students) experience in an emergency department (ED) setting, given different degrees of involvement in model building, reuse and experimentation time. Participants are explicitly tasked with learning how to increase the proportion of patients meeting the four hour wait time target within UK EDs and also satisfying their own definitions of effective performance (typically aiming for resource utilisation to be close to 100%). Measurement focuses on single-loop learning: participants' learning of strategies to meet these objectives within the ED; their attitudes towards the management of resource utilisation; and their choice of variables in experimentation.

The paper is structured as follows. Firstly, the learning themes from the simulation literature are briefly summarised and classified using formal definitions of learning taken from Argyris and Schön's (1996) theory of single and double-loop learning. Secondly, the design, materials and predictions for

the experiment are detailed. After presenting the results there is a discussion of the possible learning mechanisms in the experiment and how evaluation studies might incorporate the results.

2 Simulation and Learning

This section provides a discussion of the conceptual and theoretical background for understanding our experiment and its results. This begins with a discussion of the overarching high involvement hypothesis and how it is typically expressed in the simulation literature. This is followed by an overview of Argyris and Schön's framework of single and double-loop learning and a review of studies of learning from simulation and the complementary field of behavioural operations.

2.1 The high involvement hypothesis

Detailed studies of client learning in DES and the wider simulation community are relatively rare compared to publications on models and their results. Of those that do tackle learning and practice, discussions often refer to the hypothesis that involving the clients in model building provides much of the learning useful for aiding decisions (Alessi, 2000; Andersen, Richardson, & Vennix, 1997; Paich & Sterman, 1993; Robinson, 2004; Rouwette, et al., 2011; Rouwette, Vennix, & Mullemkom, 2002; Thomke, 1998; Ward, 1989). The outcomes of client involvement in model building might take an anecdotal form in general discussion; for example, 50% of benefit of modelling is gained simply by building the model with client involvement (Robinson, 1994). More usefully, it might take a more testable form by referring to specific learning outcomes, such as those listed in Table 1.

Table 1: Formulations of the high involvement hypothesis

- Implicit mental models of the system converted into explicit mental models (Andersen, et al., 1997)
- Refined mental models for managing the system (Rouwette, et al., 2011; Thomke, 1998)
- Improved creativity and generating new ideas for improving system performance (Robinson, 2004);
- Abstraction of general principles from models that can be transferred elsewhere (Alessi, 2000; Lane, 1994; Thomke, 1998)

Common to all of the formulations listed in Table 1 is the theory that a simulation client has a simple predictive mental model of how the system under study behaves. Involvement in model building is hypothesised to aid clients to recognise their own implicit assumptions (Andersen et al., 1997), refine and change mental models (Rouwette et al., 2007; Thomke, 1998), enhance creativity in

problem solving (Robinson, 2004), and generalise knowledge so that it can be transferred to other similar problems (Alessi, 2000; Lane, 1994; Thomke, 1998).

To illustrate the high involvement hypothesis using these definitions, consider a manager in an ED that aims to simultaneously reduce patient waiting times and increase the utilisation of ED resources to, her definition of optimality, 100%. Implicitly the manager believes that achieving these aims is just a question of resources working hard(er) to meet the targets and hence does not recognise any trade-off between the two objectives. Under the assumptions of the high involvement hypothesis the manager, through her involvement in model building, would recognise the limitations of her mental model and refine it. One way to conceptualise this refinement is as the change in an individual's attitude(s) towards controllable variables (Thomke, 1998) or competing implementation options (Rouwette, et al., 2011). For example, our ED manager may now realise that she should consider the trade-off between utilisation and waiting times when making resource decisions. If the manager has learnt correctly her attitude towards 100% utilisation of resources will decrease in strength while her attitude towards allowing a reduction in resource utilisation to achieve lower waiting times will have increased.

2.2 Single and Double-Loop Learning

A well-known framework for learning is Argyris and Schön's (1996) theory of single and double-loop learning. The starting point to understanding the framework is to assume that an individual's mental model comprises a set of variables that govern what to do or how to act in relevant situations. These governing variables constrain the actions (or decisions) individuals will take given a particular situation.

Argyris and Schön's empirical work illustrates that most individuals will attempt to keep their governing variables within acceptable limits. For example, assume our healthcare manager finds that her new ED management policies are achieving 95% utilisation of nurses, but very long waiting times in ED. As our manager's expectations have not been met, her attitude towards the management policy will be less favourable and she will attempt to find new policies that do keep governing variables within acceptable limits. Learning of this type is called *single*-loop learning: a change in attitudes towards various management actions to keep governing variables within acceptable limits. When a mismatch in expectations prompts the manager to reflect on her own mental model she undertakes a *double* learning loop: a change in governing variables and a change in actions to keep them in acceptable limits.

The learning framework set out by Argyris and Schön's framework applies not just to business management situations, but also the approach to learning or more formally the *learning systems* individuals employ. Bakken et al's (1994) experiment, using System Dynamics models, illustrates the impact of learning systems. In the experiment students and managers (executive MBAs) used a training simulation model, set in the same domain as the managers worked, followed by a second model with the same underlying behaviour but set in a different domain. All participants had to achieve a high profit with both models. In the training game, the managers followed the management approaches they used in real life. The students, having no experience of the real world system, used many alternative approaches and were rewarded with many negative (bankruptcy) as well as positive outcomes. Surprisingly, the students substantially outperformed managers in the second model.

To explain this result, the difference in learning systems between the managers and students must be examined. The governing variable the managers were attempting to satisfy was 'maximise winning and minimise losing'. Indeed they did find strategies that achieved this outcome in the training model and hence were constrained within a single-loop learning system. This is not necessarily a problem; however, it did mean that the managers failed to grasp the deeper transferable knowledge about system structure. Double-loop learning systems involve meta-learning where individuals must reflect on and correct their governing variables for learning. For example, the managers could have reassessed their need to maximise winning and adopt an approach similar to the students. To do this they would need to firstly recognise their confirmation bias and overcome any reluctance to produce negative results.

2.3   Studies that have investigated learning from models

The majority of studies that have explored learning from models have done so from an experimentation perspective using simulation gaming (e.g. Bakken, et al., 1994; Bell & O'Keefe, 1995; Langley & Morecroft, 2004; Paich & Sterman, 1993; Rouwette, Größler, & Vennix, 2004; Sterman, 1989). These have explored aspects such as transfer of learning from one game to the next (Bakken, et al., 1994), compared Visual Interactive Simulation (VIS) with traditional statistical analysis (Bell & O'Keefe, 1995) and identified factors such as model transparency that are important for learning (Rouwette, et al., 2004).

Unfortunately the evidence of successful learning from experimentation and gaming is mixed (Lane, 1995; Neuhauser, 1976; Rouwette, et al., 2004; Van der Zee & Slomp, 2009) and provides little insight into how clients learn in a model reuse project. One simple argument is that model reuse may actually aid learning, as time saved can be used to run more experimentation and more VIS with clients. More VIS, for example, would be beneficial as it seemingly aids discovery, clarification and change of clients' views and ideas about system management (Belton & Elder, 1994). Moreover, efficient experimental designs provide detailed information to clients on controllable variables and competing implementation options (Law, 2007). The other side of this argument is that client learning may be affected by *not invented here syndrome* (Pidd, 2002; Robinson, et al., 2004), that is, lack of stakeholder trust in the model, in which case at least some project time needs to be allocated to familiarising the clients with a reused model and building credibility.

Outside of gaming only a small number of rigorous studies exist that explore client learning in practice (Rouwette, et al., 2011; Thomke, 1998). Although these studies demonstrate a general change in simulation client attitudes towards implementation options (Rouwette, et al., 2011) and sudden changes in client attitudes towards controllable variables (Thomke, 1998), these studies do not separately analyse model building and experimentation and provide little insight into and evidence of the high involvement hypothesis.

Further to the individual of level of learning, it is worth briefly mentioning that outcomes from involvement in model building have also been studied and evaluated at the group level (Andersen, et al., 1997; Franco & Lord, 2010; Franco & Montibeller, 2010; Rouwette, et al., 2002). Evaluation at the group level refers primarily to the effect involvement has on the interaction between group members and group project performance. Key contributions are that a modelling framework provides a shared language to improve communication within a group (Franco & Lord, 2010); social construction of a model aids management teams make sense of the organisation in which they are working (Andersen, Vennix, Richardson, & Rouwette, 2007); and the degree to which a group understands system dynamics is a predictor of perceptions of information sharing quality and psychological safety (Bendoly, 2013).

2.4   Behavioural OM and OR

In recent years there has been much interest in behavioural operations management (Bendoly, Croson, Goncalves, & Schultz, 2010). This field re-examines the traditional assumptions built into operational models, such as 'equal worker skills' on a production line or 'managers following optimal

ordering policies' for inventory management, and studies how this affects operational performance. A common approach to capture this behavioural information is a controlled experiment set in a specific context (Bendoly, Donohue, & Schultz, 2006). For instance, there has been a stream of work investigating the behaviour of boundedly rational decision-makers in the supply chain context (Dimitriou, Robinson, & Kotiadis, 2009; Kalkanci, Chen, & Erhun, 2011; Katok & Wu, 2009; Loch & Wu, 2008; Su, 2008). In general this work asks students to act as suppliers or retailers working in a newsvendor setting under a different contractual arrangements e.g. the wholesale price contract, buyback contract or revenue sharing contract. Observations in a laboratory setting demonstrate that these human actors behave very differently to a rational-optimising decision-maker, such as is assumed by analytical models. Indeed, there are examples where humans perform better than expected from the assumptions of analytical models e.g. Katok and Wu (2009), Dimitriou et al (2009).

Our work has a similar interest in observing human actors in a decision-making environment, but our focus is on the use of models to aid learning and decision-making rather than on the decision-making itself. This places our work in the field of behavioural operational research which 'focuses on the psychological and social interaction-related aspects of model use in problem solving' (Hämälläinen, Luoma, & Saarinen, 2013).

2.5 Conclusions about Simulation and Learning

Although the high involvement hypothesis appears credible there are few empirical studies providing evidence of its existence. Given the difficulties of exploring learning in simulation practice, there is relatively little insight into the mechanisms that benefit client learning when they are closely involved in model building or understanding of how model reuse may alter the learning approach. This lack of knowledge prevents a focussed exploration of learning in practice and further development of the theory and evidence base for the effectiveness of simulation as a decision aid. The remainder of this paper describes a laboratory experiment to address this area. The experiment tests single-loop learning within studies where a participant is involved in different levels of model building, experimentation and reuse. Discussion of results also considers the participants use of single and double-loop learning systems.

3 Experimental Design and Predictions

Observation and measurement of client learning from modelling is difficult in practice, as it is often uncertain if other factors affected learning or if the same learning would have been achieved from a different decision making approach. An experimental approach such as ours provides the opportunity to study learning within a controlled and simplified environment and provides insight into where evaluation of learning in fieldwork might best be focussed. We sought to answer three research questions:

Q1. Can the high involvement hypothesis be demonstrated empirically?

Q2. What mechanisms aid client learning from involvement in model building and reuse?

Q3. Is there any evidence that single-loop learning systems interact with learning from DES models?

This section details a methodology for answering these questions using a laboratory experiment. Details are presented for the independent variables, dependent variables, predictions, participants, experimental materials and procedure.

3.1 Independent Variables

To answer the research questions the experiment manipulates one independent variable (IV): the simulation study process. We include three levels of IV, as illustrated in Figure 1. The first two levels provide a comparison of model building and model reuse using a fixed budget of time, labelled as model building with limited experimentation (MBL) and model reuse (MR) respectively. Experimentation is limited in MBL over MR as more time is required to build the model. Superior performance in MBL over MR would provide evidence of the high involvement hypothesis in action. To explore the interaction of experimentation with the learning mechanisms of model building we include a third condition with an extended time budget labelled as MB. In this condition participants spend as much time on experimentation as in MR, but they are also involved in model building as in the MBL condition. A comparison between MR and MB provides the opportunity to study how involvement in model building influences the participant's behaviour in experimentation.
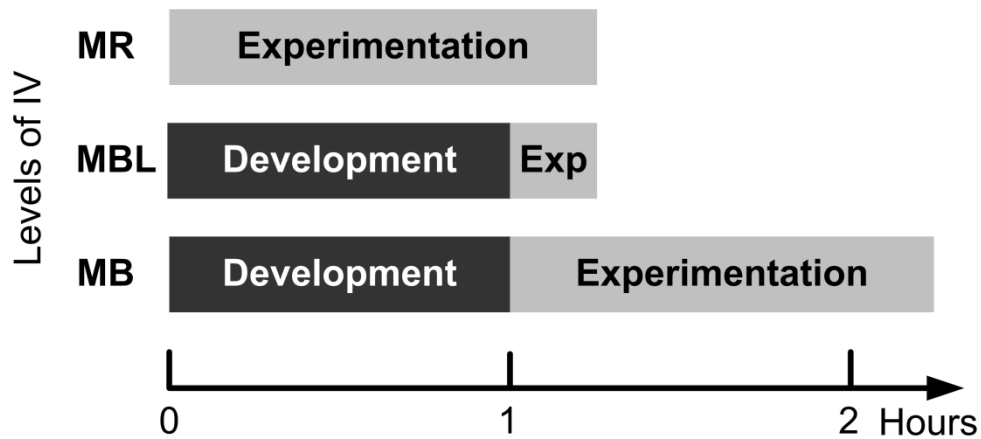
Figure 1: Differences in the duration of activities between the three conditions

3.2 Dependent Variables

There are two measures of learning within the experiment: the participant's selection of variables in experimentation and the change in a participant's attitude to managing resource utilisation in the ED. We chose resource utilisation as the focus of our measurement because a common task of a DES modeller is to aid clients' understanding of the trade-off between resource utilisation and time in the system. Other research has also indicated that experienced managers do not always appreciate the trade-off, for instance in manufacturing (Suri, 1998).

We operationalised measurement of single-loop learning by using the attitude measurement procedures of the Theory of Planned Behaviour (Ajzen, 1991): a well-known and substantiated theory (see Ajzen, 2001) from social psychology that has been used elsewhere in the simulation literature (Rouwette, et al., 2011). Two attitude variables were developed during a pilot phase. Firstly, *MaxUtil* that represents the attitude towards attempting to achieve 100% utilisation of a resource. Secondly, *TradeUtil* that represents the attitude towards allowing some spare capacity, on average, to improve performance (i.e. time in system).

The two attitudes are measured pre- and post-test using identical questionnaires (see supplementary material). The participant reads a series of statements in the form of beliefs about the advantages or disadvantages of management actions at the hospital. Participants then rate the likelihood that a belief is true (b) on a seven point scale (1= Extremely Unlikely; 7 = Extremely Likely), and how desirable the outcome of this belief is (e) on a bipolar seven point scale (-3 = Extremely Bad, +3 = Extremely Good). The measure of an attitude (A_i) is constructed from the summation of the *j* products of the subjective probability and outcome evaluations (Ajzen, 1991). This is summarised in equation 3.1 where k is the number of attitude variables.

$$A_i = \sum_{j-1}^{k} b_j e_j \qquad\qquad (3.1)$$

MaxUtil is constructed from three beliefs and *TradeUtil* from five, giving scale ranges of ±63 and ±105 respectively. Single-loop learning is inferred from a reduction in the strength of *MaxUtil* and an increase in the strength of *TradeUtil*.

3.3 Hypotheses

The experiment tests three hypotheses to answer Q1; these are summarised in Table 2. Firstly, if the high involvement hypothesis holds true then MB participants should show more creativity in experimentation than MR participants. We test this by comparing the number of new variables MB and MR select for analysis (hypothesis 1). Secondly, participants in the model building (MB and MBL) change will experience more attitude change towards queue management than MR (hypothesis 2). Finally, we analyse a subgroup where participants involved in model building (MB and MBL) should typically have more attitude change in the correct direction than MR. Note that in a real simulation study one might argue that a decision maker is neither right nor wrong, but instead makes a decision based on his/her own worldview. In the simplified world of the laboratory experiment it is possible to determine the direction of attitude change that improves the performance of a system and the direction of attitude change that does not improve performance. For the sake of clarity and understanding these directions will be labelled as 'the correct direction' (hypothesis 3) and the 'incorrect direction' of attitude change.

For *MaxUtil* the correct direction of attitude change is represented by a decrease in score. I.e. participants will no longer seek to work resources at their maximum utilisation. In contrast, high scores on *TradeUtil* represent a favourable attitude towards allowing spare capacity in order to achieve lower queuing times. Therefore the correct direction subgroup is identified by an increase in attitude strength.

Table 2: Single-Loop Learning Hypotheses

Hyp	Measure(s)	Prediction
1	Experimentation with new variables	MB > MR
2	Attitude change	MB and MBL > MR
3	Attitude change in the correct direction	MB and MBL < MR

3.4   Participants

Sixty four business undergraduates volunteered to take part in the research (first year undergraduates = 41[64%], second year undergraduates= 23[36%], male = 35[55%] female=29[45%], age range = 18-22). All students were registered on modules taught by Warwick Business School, but none had prior experience of simulation. After sign-up was completed students were anonymised and randomised to a condition. To encourage participation the students were paid a small fee for their time. To improve participant motivation an additional cash prize was available for the best performance.   Three participants were excluded from the analysis due to unusual multivariate profiles. In particular, these participants demonstrated good understanding of the queuing concepts tested in the experiment during a post-test interview, but provided contradictory and extreme scores all attitude measures.

3.5   Materials

3.5.1   Emergency Department Problem

The experiment is based around a case study problem. This takes the form of a fictional Emergency Department (ED) - St. Specific's. The objective is to reduce the percentage of patients that are in ED for longer than four hours over a six month time horizon as well as achieve participants' own definitions of effective performance, such as acceptable resource utilisation targets. Embedded in the case are several contrived learning outcomes for participants. For example, a participant may discover:

- Maximising resource utilisation of doctors and nurses over the six month period leads to a large number of performance target breaches.
- There is always a trade-off between the mean time patients spend in the ED and the utilisation of ED resources.

Participants are provided with a default set of scenarios to investigate; for example, the reallocation of nurses across shifts. Participants in the MB and MR conditions are also able to suggest new scenarios which may investigate the default scenarios further or something else entirely. For example, a participant may investigate the effect of patient prioritisation on performance.

3.5.2   Simulation Models

The case study is based around a simplified version of the generic ED models described in Fletcher et al. (2007) and Günal and Pidd (2011). The model is implemented in Simul8 Education Edition 2009

(Simul8, 2009) and the same model is used by all participants during experimentation. We also created 11 sub-model versions of the simulation that are tested and explored by MB and MBL participants during their involvement in model building.

3.6 Procedure

The procedure followed by the three conditions is illustrated in Figure 2. All participants initially fill in the attitude questionnaire and then receive an introduction to simulation: a researcher follows a script to explain the SIMUL8 software, the visual display, output measures, multiple replications and what-if scenario analysis. After the introduction the procedure is manipulated depending on the condition that the participant is allocated. On completion of the condition each participant fills in a post-test questionnaire.
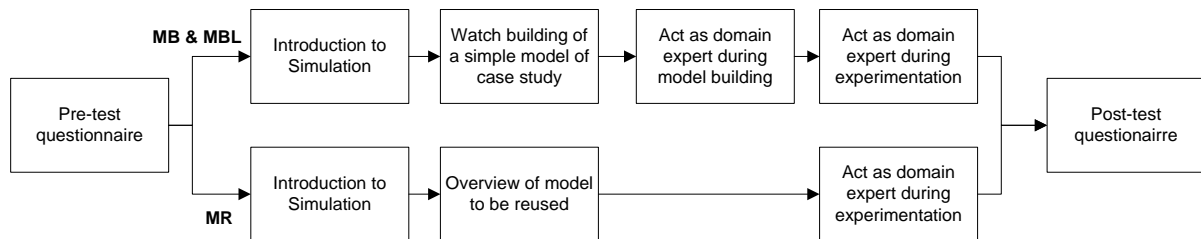

Figure 2: Differences in procedure between the three conditions

3.6.1 Model Building Procedure

Participants sit with the researcher in front of a desktop personal computer loaded with Simul8 and Microsoft Excel 2007. The researcher modifies the simulation model during building and experimentation at the request of the participant. Results from the model are automatically loaded into a Microsoft Excel Spreadsheet for participants to review as they wish.

Participants in MB and MBL go through five rounds of model building and validation. In each round participants are presented with a visual model they can watch or step through, a results screen and a single A4 page detailing the change to the model, simplifications included, and assumptions. Participants review this information and decide if the model is fit for purpose or suggest how it should be changed (by progressively adding more detail) to better represent the case study. In fact, the combinations and order of changes that a participant can request are limited; hence a number of

pre-constructed sub-models (11) are available to the researcher. Once the request for change is made the researcher simply opens up the appropriate sub-model.

Participants can ask to remove any of the simplifications or clarify their meaning. Table 3 provides an excerpt from the first conceptual model that a participant sees and is useful for explaining this process. The first simplification in Table 3 is an example of a simplification that cannot be removed. If a participant asks for an explanation of the simplification or for it to be removed, the researcher provides a scripted reason for why it is included. In the case of the first simplification listed, no data is available to model the movement of nurses in greater detail and it is difficult to make any sensible assumptions. Participants can ask for repeats of this explanation at any stage in the experiment. The second simplification is an example of a where the model is oversimplified. Participants need to ask to remove it in order to reach the final model.

Table 3: Excerpt from stage one conceptual model

Simplifications
- Nurse multi-tasking is not modelled in detail. For example, if a nurse can do two things at once then two nurse slots are included in the model;
- An average number of nurses are included. Staff shifts are not explicitly modelled;

Assumptions
- Nurses always correctly evaluate the level of emergency a patient represents.

3.6.2 Model Reuse Procedure

Participants in MR undergo the same simulation education as MB and MBL. However, MR participants are informed that a model was developed for a different hospital with a similar process and objective. They are then presented with the full computer model (the final model reached by MB and MBL), the conceptual model and results from a batch run of the model. The following procedure is then followed:

1. The model is run in visual interactive mode and participants are talked through the process; for example arrivals, priority queuing, emergency treatment and visiting radiology;
2. The explanation for the result screen given to MB and MBL at stage one of model building is replicated;
3. Participants are presented with a list of assumptions and simplifications in the model.

Participants are reminded that the model was developed for another similar hospital. Thus they must assess the model's fitness for purpose. They are allowed to ask questions about model logic, results, simplifications and assumptions. Once participants are satisfied that suitable verification and validation (V&V) has been completed they proceed to experimentation.

3.6.3 Experimentation Procedure

MB and MR participants must investigate six scenarios predefined in the case study and MBL is limited to investigating three predefined scenarios. All scenarios relate directly to the two attitudes measures. MB and MR participants are also free to suggest new scenarios based on their own ideas. These new scenarios may be directly relevant to the measures of the experiment, i.e. resource based, or may be something entirely new. Participants are provided with a summary of scenario results is in Microsoft Excel and they are able to drill into the results in more detail.

Before analysis of every scenario, participants are reminded that, if desired, they can watch the model's visual display as well as review batch run results. Additionally, once results are loaded into the scenario comparison spreadsheet, participants are reminded that they can drill further into the data if desired.

4   Results

4.1   Hypothesis 1: Experimentation with new variables (MB>MR)

4.1.1   Number of variables

The majority of MB participants (19 out of 20) identify and run experiments using new variables (median$_{MB}$ = 2.0 new variables identified per participant) compared to MR (median$_{MB}$ = 0.0 new variables identified per participant). Among the MR participants new variables are explored by less than 50% (9 out of 20) of participants (median difference = 2.0; $p < .05$; $r = .36$); supporting hypothesis 1.

In total MB participants typically simulate 3.5 scenarios more than MR (median$_{MR}$ = 10.0, median$_{MB}$ = 13.5, significant at $p < .1$), as MR use some of the experimentation time to learn how to use a simulation model. We checked for a learning effect during experimentation in the MB condition by analysing the first scenario where participants identified a new variable. The exploration of new variables in MB was not biased to the later scenarios with participants typically identifying a new variable within the first three to four scenarios (median = 3.5; inter-quartile range = 3.5) and 30% of participants identifying a new variable in the first scenario.

4.1.2 Choice of variables

During the experiment it was observed that MB participants appeared to choose more validation type scenarios early in the experimentation phase. By *validation type* we refer to scenarios that concern the exploration of assumptions and simplifications included in the model; for example, two participants were concerned that the nurses did not make any mistakes when triaging patients. Table 4 lists the new variables explored in scenario one. Three of these variables are listed as *validation* scenarios. The two remaining scenarios are labelled as *performance* as they are believed to improve performance. MB participants only experiment with new variables of the validation type in scenario one.

Table 4: New variables identified in the first scenario

		No. of Scenarios		
Variable	**Type**	**MR**	**MB**	**Total**
Multi-tasking*	Validation	1	3	4
Alternative Shift Patterns	Performance	1	0	1
Misdiagnosis of patients at triage	Validation	1	1	2
Removal of prioritisation system	Validation	0	3	3
Fast Jr Doctors	Performance	1	0	1

*For example, treatment of major emergency patients requires more than one doctor slot.

A similar picture is observed when all scenarios are considered. Table 5 lists new variables selected by participants along with the number of scenarios and the number of participants that explore them in MB and MR. One difference that stands out from the rest is the number of participants that ran scenarios concerning the multi-tasking of resources in the model (i.e. doctors and nurses treating multiple patients at once; see Günal and Pidd (2006) for details). This was largely a validation exercise with participants exploring the assumptions in the model.

Another example is the removal of the prioritisation system explored by five MB and zero MR participants. Although this difference is small, and hence any interpretation should be taken with caution, it was noted that these MB participants simulated these scenarios in response to noticing changes of model behaviour and results during model building. Furthermore, a number of other MB participants noticed the same behaviour, but were unable to pinpoint the prioritisation system as the cause of the behaviour (for example participants MB1 and MB20 simply watched the model running for an extended period).

In summary the choice of variables for experimentation supports hypothesis 1: MB participants identified a greater variety of variables than MR participants. In addition validation of the model tended to prompt MB participants to choose related variables during experimentation.

Table 5: New variables and their use in experimentation

Variable	No. of Scenarios		No. of participants	
	MR	MB	MR	MB
Multi-tasking	1	11	1	10
Alternate shift patterns	7	8	5	7
Misdiagnosis of patients in triage	1	2	1	1
Removal of prioritisation system	0	5	0	5
Treatment times	2	4	1	3
Dedicated Resources	0	4	0	3
X-ray rooms resources	3	3	3	3
Extra evaluation room resources	0	1	0	1
Pooling resources	0	1	0	1
Registration Desks	0	1	0	1
Better x-ray decision making	3	0	3	0
Cubicle sharing	1	0	1	0
Fewer patients brought to hospital	2	0	1	0

4.2 Hypotheses 2 and 3: Attitude Change

Attitude change focussed on participants' attitudes towards managing resource utilisation in the ED. Pre-test attitudes were similar across the three conditions for *MaxUtil* (overall median = 30; p > .1) and *TradeUtil* (overall median = 23; p > .1). We report the median change from pre-to post-test for these measures, labelled as *ΔMaxUtil* and *ΔTradeUtil,* respectively*,* along any significant conclusions (α = .1) and a standardised measure of effect size. Use of a correlation coefficient *r* as a standardised measure of effect size is common in social psychology and other empirical sciences and allows for comparison of effects between measures and across studies (Cohen, 1989; Field, 2009).

4.2.1 Hypothesis 2: Overall results
The overall attitude change results show no significant effects between conditions and do not support hypothesis 2 (p > .1). We report medians and inter-quartile ranges for each condition in Table 6.

For *ΔMaxUtil* the overall result median changes are similar across the conditions while MB results have higher variability (Inter-quartile range$_{MB}$ = 31.0) than those for MBL and MR (Inter-quartile range$_{MBL}$ = 17.5; Inter-quartile range$_{MR}$ = 17.3).

Median differences are again similar for *ΔTradeUtil* across the three conditions. The MBL results are less variable (inter-quartile range$_{MBL}$ = 17.5) than those for MB and MR (inter-quartile range$_{MB}$ = 38.3; inter-quartile range$_{MR}$ = 35.3).

Table 6: Overall attitude change results

	Attitude Change Median (IQR)			
Measure	MB	MBL	MR	Conclusion(s)
ΔMaxUtil	-7.0 (31.0)	-6.0 (17.5)	-9.5 (17.3)	No difference
ΔTradeUtil	7.5 (38.3)	10.0 (17.5)	11.0 (35.3)	No difference

4.2.2   Hypothesis 3: Correct Direction (MB and MBL > MR)

The majority of participants experienced attitude change in the correct direction on both measures (65%-80%, Table 6). For *ΔMaxUtil,* MB participants experience more attitude change than MR and MBL (with greater variation in outcome); supporting hypothesis 3. However, inference results, in Table 7 only support the view that MB > MR (r = .20, p < .1). This is because the MB versus MBL test is two-tailed, as we had no specific prediction for the comparison, leading to lower power. Note, however, that the effect size for 'MB versus MBL' is similar to 'MB versus MR' (r = .27; p > .1).

MBL experience less *ΔTradeUtil* in the correct direction than MR and MB with less variation in the outcome (Inter-quartile range$_{MBL}$ = 13.0; Inter-quartile range$_{MB}$ = 24.5). In both instances this is a medium effect size (r = .30 to .40, *p* < .1). Thus while the results for *ΔMaxUtil* provide some support for hypothesis 2 the results for *ΔTradeUtil* do not*.*

Table 7: Attitude Change Results (Correct Direction Subgroup)

	% Participants			Attitude Change Mdn (IQR)			
Measure	MB	MBL	MR	MB	MBL	MR	Conclusion(s)
ΔMaxUtil	70	67	65	-21.5 (24.3)	-10.0 (8.3)	-13.0 (8.0)	MB > MR
ΔTradeUtil	70	81	70	23.0 (24.5)	15.0 (13.0)	22.5 (25.8)	MB = MR > MBL

4.2.3   Incorrect Direction Subgroup

To explore if DES users can misunderstand model results, we also include results for a second subgroup: the incorrect direction of attitude change. The sample size for the incorrect direction of

attitude change is very small (n ≈ 6 per condition); hence we provide a graphical analysis only. Figure 3 presents quantile-quantile (Q-Q) plots for *ΔTradeUtil* for the 'MR versus MBL' and 'MB versus MBL' comparisons. The points below zero on either axis represent quantiles of attitude change in the incorrect direction of *ΔTradeUtil*. In both Q-Q plots the quantiles less than zero are beneath the line y=x, indicating that there was a general trend for MBL to experience less *ΔTradeUtil* in the incorrect direction. Hence the data illustrate that a minority of MB participants could still misunderstand the model's results.

MR versus MBL   MB versus MBL
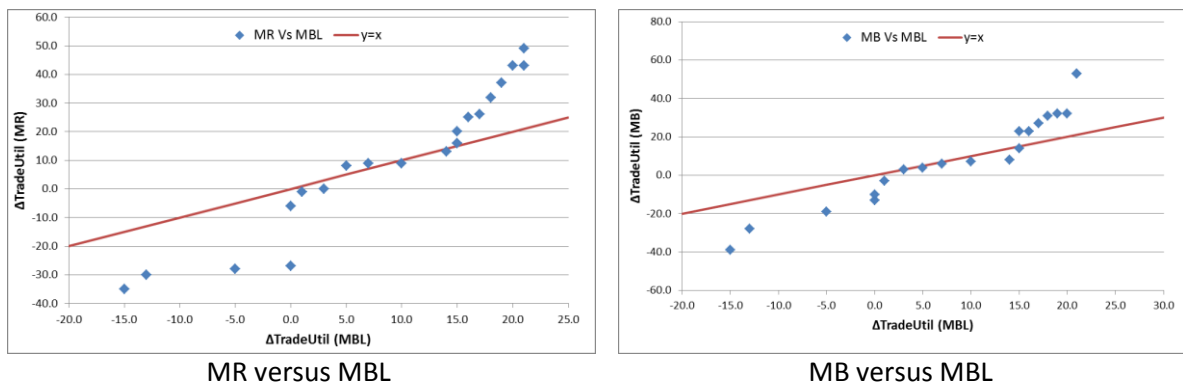
Figure 3: Q-Q plots of ΔTradeUtil. If the points lie on the line y=x then the quantiles of the two distributions are equal. If the points lie above the line y=x then the quantiles of the condition represented by the y axis are greater than the condition represented by the x-axis. If the points lie below the line y=x then the quantiles of the condition represented by the y axis are less than the condition represented by the x-axis.

5 Discussion

The results of the experiment provide some evidence that support the high involvement hypothesis; however, there were also some unexpected results. This section discusses the results in the context of the research questions and summarises the key limitations of the analysis.

5.1 Can the high involvement hypothesis be demonstrated empirically?

Our first research question asks if the high involvement hypothesis can be demonstrated empirically. In practice it is often difficult to deduce if, how and why clients learn during a DES study; hence it is often not possible to provide any evidence supporting the high involvement hypothesis beyond a retrospective and anecdotal discussion. One of the strengths of this studys' approach is that if the high involvement hypothesis is real and substantial, it should be demonstrable in the simplified environment of the laboratory experiment.

Our results demonstrate that client learning is not as straightforward as is often portrayed. The experimentation results support the hypothesis, as involvement in model building prompted participants to explore problem variables that were not predefined for them. Furthermore, the exploratory analysis of variable choice suggests that MB participants developed a good understanding of the model and were ready to explore some *validation type* factors during experimentation. However, the results for attitude change are not as clear cut. In particular, model building participants only outperformed model reusers when experimentation time was not limited.

5.2 What mechanisms aid client learning from involvement in model building and reuse?

To explain the contradictions in the results, our second research question asks what mechanisms aid client learning from involvement in model building and model reuse? Here we propose three learning mechanisms summarised in Table 8 along with corresponding key results from the experiment.

Result	Proposed Mechanism
MB participants identify and explore more new variables in experimentation than MR participants;	The scope of participant's available options benefitted from the incremental development and validation of a simple model to a relatively complex one.
MB participants had the highest Δ*MaxUtil* in the correct direction;	An interaction between complexity of the concept being studied and attitude strength meant that MB benefitted from more time.
MBL participants experienced the least Δ*TradeUtil* in the correct direction;	Learning about the relationship between resource utilisation benefits from the classic Kolb learning cycle.

5.2.1 Mechanism 1: Incremental Validation

One explanation for MB participants identifying and choosing to experiment with more new and validation type variables than MR is that the participant's understanding was influenced by the change in model output(s) as the level of detail was gradually increased. One such example is the effect of patient prioritisation and variable inter-arrival times on the time spent in the ED. In early versions of the model patient prioritisation is not included and, due to the length of treatment, major emergency patients spend longer in the ED than minor patients. Once patient prioritisation is added this reverses. Participant MB19 provides an example of a common response to this result:

"This [the proportions of minor and major emergency target breaches] is not very realistic. Of course, it takes a longer time to treat major emergencies than the minor ones. So what is the catch? Where is the mistake in the model?"

Participant MB19 then spent an extended time in V&V; watching the model run, pausing it and looking at the contents of queues proved to be useful. This is unsurprising given the stated benefits of Visual Interactive Simulation (Belton and Elder, 1994). However, visual inspection was not always enough; some participants also chose to simulate the system without prioritisation included to see the effect on performance. Of course, the effect of patient prioritisation could be questioned by model reusers as well, but it was only model builders who were surprised and altered their behaviour. As novel and surprising events are remembered better and increase attention (Butterfield & Metcalfe, 2006; Tulving & Kroll, 1995) learning opportunities should be greater in these participants. This results is also an indication that participants are using a double-loop

learning system during model building and are open to question their own assumptions about system behaviour.

One concern, however, is when experimentation time is limited the importance of the variable might just be achieved by following a heuristic; for example, the participant applies more weight to the importance of a variable because it is novel to them. Hence, although involvement in modelling appeared to generate more ideas in the experiment it may also have introduced a cognitive bias that hinders further learning. For example, individuals may over rely on a specific piece of information (see anchoring in Tversky & Kahneman 1974) and neglect information from other areas of the model.

5.2.2 Mechanism 2: Attitude strength and complexity interaction

It has been shown elsewhere that managers struggle to appreciate the relationship between resource utilisation and system time in queuing systems (Suri, 1998) and in our study pre-test attitudes were typically strong (median = 30). Therefore a simple explanation for MB experiencing the most $\Delta MaxUtil$ is that that they had more time than MR and MBL. For example, in a debrief after the experiment participant MB9 pointed out that he noticed "during model building" that 100% resource utilisation was not working and that it was "surprising" that a lower utilisation, from obtaining more resource, helped. MB9 then had plenty of time to test and re-test this surprising finding during experimentation. This would be more difficult in MBL, where experimentation was controlled, and MR, where participants would only first observe such behaviour during experimentation and have less time to think it through.

5.2.3 Mechanism 3: The classic learning cycle and experimentation

Hypotheses 2 and 3 did not hold true in the MBL condition. Indeed as the conditions that contained more experimentation (MB and MR) generally had a larger $\Delta TradeUtil$ in the correct direction along with some graphical indication of smaller $\Delta TradeUtil$ in the incorrect direction. A plausible conclusion is that the classic 'Kolb' learning cycle (Kolb, 1984)- conceptualise, test, observe and reflect - aids learning about resource utilisation. Hence, restricted experimentation inhibits learning about resource utilisation. Of course, other areas of learning might benefit more from involvement in model building than experimentation and other factors such as model complexity may interact; however, the example does illustrate that it is not as clear cut as the high involvement hypothesis would imply.

5.3 Is there any evidence that single-loop learning systems interact with learning from DES models?

Our third research question asks if there is any evidence that single-loop learning systems interact with learning from DES models. The $\Delta TradeUtil$ result for MB indicates that even when a participant is given plenty of time and involved in model building (s)he can still substantially misunderstand the relationship between resource utilisation and performance. This result may in part be down to the dominance of a single-loop learning system in these individuals and an inability to learn how to learn (a double-loop system). In single-loop learning systems individuals seek to confirm they are correct (maximise winning) about managing utilisation as opposed to testing that knowledge and possibly receive negative results (minimise losing). For example, participant MB14 ($\Delta TradeUtil$ = -20) quickly disregarded results showing reduction in performance from a self-choice resource reallocation scenario. Appearing to be quite embarrassed that her choice did not produce positive results, MB14 quickly moved onto something more obvious (adding extra resources) that gave her an improvement in performance. MB14 then continued to choose this type of scenario only. This behaviour is a form of 'confirmation bias' (Bell & O'Keefe, 1995; Fraser, Smith, & Smith Jr, 1992). We note that this behaviour occurred in the minority of cases. A possible explanation is related to the link between understanding of system dynamics and psychological safety (Bendoly, 2013). In particular, as most participants built up an understanding of the dynamics of the model then this reduced the perceived risk of a negative response (particularly looking foolish) in front of the researcher. This helps explain the effective learning seen in most participants.

5.4 Limitations and Further Work

There are, of course, some limitations with a laboratory study of this nature. We now discuss these limitations as well as opportunities for further work.

As students were used as participants, some care should be taken with interpretation of the results summarised in Table 8. Particularly as real world managers are less likely to listen to advise (Yaniv, 2004) or may have political aims that affect how a model is used. Nevertheless, as the experiment gave the novice simulation clients every opportunity to learn how to manage the case study ED more effectively, failures to meet expected outcomes may indicate that managers are even less likely to do so.

The sample size was a constraint on the research due to the funds available and the time required to run the individual experiments. To aid the comparison of the attitude results to further similar

research we provide standardised effect sizes; enabling some judgement on the size of learning differences beyond statistical significance between groups. Ultimately, however, more confidence in the results could be gained by employing a larger sample size and stricter alpha level. For attitude measures power analysis suggests that a sample size of around 40 participants in each condition (i.e. a total of 80 in a comparison and a total sample size of 120) resulting in a total of 330 hours laboratory time. Effect sizes for differences in selection of new variables, however, were larger and would stand up to stricter significance testing; replication of these should be possible with group sizes of less than 40.

We attempted to improve recruitment rates and participant motivation in the task by providing a cash prize for the best performance in addition to a participation fee. A weakness in this design is the potential for 'reward-effort' bias. That is, if a participant believes that (s)he has no chance of receiving a reward then (s)he may have less task motivation or give up on the task entirely. The intrinsic motivation of participants is often measured as the degree to which participants return and persist at a task during a free choice period after the experimental phase (Deci, Koestner, & Ryan, 1999). Given the length of the experiments, we chose not provide a free choice period post-experiment and cannot substantiate that a 'reward-effort' bias was or was not present. Instead we provide two propositions as to why the reward may not have substantially been effecting motivation. Firstly, many participants in the MBL condition (who could only run three scenarios) were keen to continue with the task post-experiment. Indeed the researcher was either asked to provide feedback on other ideas participants wished to explore at the end of the experiment or via e-mail contact at a later date. Secondly, many participants asked the researcher to explain or validate their understanding of the trade-off between resource utilisation and queuing time during the post-experiment debrief. We propose that, as the initial attitude participants held towards maximising resource utilisation were strong, the surprise they experienced when the model behaved in an unexpected way increased interest and helped negate their 'reward-effort' bias.

A further limitation is that the experiment only used a single model. The choice of context may be important due to possible differences in objectives. In the case study, as in real life, the participants faced a challenging time based target coupled with high variation in arrivals and the need for value for money from resources. This high service level requirement meant that the resource utilisation 'problem' (i.e. its trade-off with queuing time) was a significant challenge for the participants. A model from a different context may not have an objective that puts them at odds with the resource

utilisation problem in such an aggressive manner. Thus the participant may be able to focus on other aspects of the problem - aspects in which experimentation is not as important as model building.

Further work should use the results of this study to plan what variables should be considered to evaluate client learning in fieldwork studies of DES projects. We make two recommendations. Firstly, we note that involvement in model building prompted participants to explore particular areas of uncertainty during experimentation in the form of validation type scenarios. In evaluation studies it would seem important to keep a longitudinal record of variables explored and for what purpose: validation or improvement. In the experiment the modeller's role was scripted and all ideas came from the participant. In a fieldwork study, the source of the idea, either client or modeller, should also be objectively tracked by a third party. At the simplest level this might be a comparison of the variables the client believes are important at project commencement, after model building and after study completion. An alternative more detailed approach might make use of verbal protocol analysis (Tako & Robinson, 2010) during client and modeller meetings.

Secondly, DES evaluation studies should keep close track of how and if clients learn about the relationship between resource utilisation and queuing. To validate our findings research questions should focus on the methods modellers find most effective to communicate the relationship to the client, for example, via interactive experimentation, reports or presentation, and if learning is dominated by confirmation bias (a single-loop system).

6 Conclusions

It is often assumed that simulation clients and decision makers experience much of their learning during involvement in model building. This study provides a simple test of this high involvement hypothesis: what happens to learning if we build and limit experimentation versus if we reuse a model. The simulation clients (students) still learnt about aspects relevant to the queuing problem (resource utilisation) when they reused a model; in fact, these participants learnt more than those that were involved in model building, but had very limited time for experimentation. However, this came at a cost of less variety in the choice of scenarios for experimentation. This latter result may be general to other modelling approaches as well i.e. involvement in modelling broadens the scope of a client's mental model of a problem. It may be less easy, however, to involve clients in model building and experimentation with some other methods, for example, analytical queuing models. Hence, this result may be specific to approaches which lend themselves to visual interactive modelling and live experimentation, such as, system dynamics and discrete-event simulation.

The results of this experiment should, of course, be subject to further empirical testing; however, they do illustrate that learning in simulation studies is not as clear cut as is often assumed. In time constrained simulation studies, practitioners may wish to consider the learning needs of their clients when deciding whether to build the model and so limit experimentation, or whether to reuse a generic model enabling more time for experimentation.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*, 179-211.

Ajzen, I. (2001). Nature and Operation of Attitudes. *Annual Review of Psychology, 52*, 27-58.

Alessi, S. (2000). Building Versus Using Simulations. In *Intergrated and Holistic Perspectives on Learning Instruction and Technology* (pp. 175-196): Kluwer Academic Publishers.

Andersen, D. F., Richardson, G. P., & Vennix, J. A. M. (1997). Group model building: adding more science to the craft. *System Dynamics Review, 13*, 187-182-181.

Andersen, D. F., Vennix, J. A. M., Richardson, G. P., & Rouwette, E. A. J. A. (2007). Group Model Building: Problem Structing, Policy Simulation and Decision Support. *The Journal of the Operational Research Society, 58*, 691-694.

Argyris, C., & Schön, D. A. (1996). *Organisational Learning II: Theory, Method and Practice*: Addison Wesley.

Bakken, B., Gould, J., & Kim, D. (1994). Experimentation in Learning Organisations: A Management Flight Simulator Approach. In *Modelling for Learning Organisations*. Portland: Productivity Press.

Bell, P. C., & O'Keefe, R. M. (1995). An experimental investigation into the efficacy of Visual Interactive Simulation. *Management Science, 41*, 1018-1038.

Belton, V., & Elder, M. (1994). Decision support systems: Learning from visual interactive modelling. *Decision Support Systems, 12*, 355-364.

Bendoly, E. (2013). System dynamics understanding in projects: Information sharing, Psycholgical Safety and Performance Effects. *Production & Operations Management, In Press*.

Bendoly, E., Croson, R., Goncalves, P., & Schultz, K. (2010). Bodies of Knowledge for Research in Behavioral Operations. *Production and Operations Management, 19*, 434-452.

Bendoly, E., Donohue, K., & Schultz, K. L. (2006). Behavior in operations management: Assessing recent findings and revisiting old assumptions. *Journal of Operations Management, 24*, 737-752.

Bowers, J., Ghattas, M., & Mould, G. (2012). Exploring alternative routes to realising the benefits of simulation in healthcare. *Journal of the Operational Research Society, 63*, 1457-1466.

Butterfield, B., & Metcalfe, J. (2006). The correction of errors commited in high confidence. *Metacognition and Learning, 1*, 69-84.

Cochran, J. K., Mackulack, G. T., & Savory, P. A. (1995). Simulation project charactoristics in industrial settings. *Interfaces*, 104-113.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examing the effects of extrinsic rewards on intrinsic motivation. *Pychological Bulletin, 125*, 627-668.

Dimitriou, S., Robinson, S., & Kotiadis, K. (2009). The impact of human decision makers' individualities on the wholesale price contract's efficiency: Simulating the newsvendor problem. In *Simulation Conference (WSC), Proceedings of the 2009 Winter* (pp. 2353-2364).

Fletcher, A., Halsall, D., Huxham, S., & Worthington, D. (2007). The DH Accident and Emergency Department model: a national generic model used locally. *Journal of the Operational Research Society, 58*, 1554-1562.

Fletcher, A., & Worthington, D. (2009). What is a `generic' hospital model? - a comparison of `generic' and `specific' hospital models of emergency patient flows. *Healthcare Management Sciencel, 12*, 374-391.

Franco, A. L., & Lord, E. (2010). Understanding multi-methodology: Evaluating the perceived impact of mixing methods for group budgetary decisions. *Omega, 39*, 362-372.

Franco, A. L., & Montibeller, G. (2010). Facilitated modelling in operational research. *European Journal of Operational Research, 205*, 489–500.

Fraser, J. M., Smith, P. J., & Smith Jr, J. W. (1992). A catalog of errors. *International Journal of Man-Machine Studies, 37*, 265-307.

Günal, M., & Pidd, M. (2006). Understanding accident and emergency department performance using simulation. In *Proceedings of the 2006 Winter Simulation Conference* (pp. 446-452). Piscataway, New Jersey.

Günal, M., & Pidd, M. (2011). DGHPSIM: Generic Simulation of Hospital Performance. *Transactions on Modeling and Simulation (TOMACS), 21*, doi:10.1145/2000494.2000496.

Hämälläinen, R. P., Luoma, J., & Saarinen, E. (2013). On the importance of behavioral operational research: The case of understanding and communicating about dynamic systems. *European Journal of Operational Research, In Press*.

Kalkanci, B., Chen, K.-Y., & Erhun, F. (2011). Contract complexity and performance under asymmetric demand information: An experimental evaluation. *Management Science, 57*, 689-704.

Katok, E., & Wu, D. Y. (2009). Contracting in supply chains: A laboratory investigation. *Management Science, 55*, 1953-1968.

Kolb, D. A. (1984). *Experiential Learning. Experience as The Source of Learning and Development*. Englewood Cliffs: Prentice Hall.

Lane, D. (1994). Modeling as Learning: A Consultancy Methodology for Enhancing Learning in Management Teams. In *Modeling for Learning Organisations*. Portland, Oregon: Productivity Press.

Lane, D. (1995). On a Resuragance of Managment Simulations and Games. *Journal of the Operational Research Society*, 604-625.

Langley, P. A., & Morecroft, J. D. W. (2004). Performance and learning in a simulation of oil industry dynamics. *European Journal of Operational Research, 155*, 715-732.

Law, A. M. (2007). *Simulation Modelling and Analysis*. Boston: McGraw-Hill International.

Loch, C. H., & Wu, D. Y. (2008). Social preferences and supply chain performance: An experimental study. *Management Science, 54*, 1835-1849.

Neuhauser, J. T. (1976). Business Games Have Failed. *The Academy of Management Review*, 124-129.

Paich, M., & Sterman, J. D. (1993). Boom, Bust and Failures to Learn in Experimental Markets. *Management Science, 39*, 1439-1458.

Pidd, M. (2002). Simulation software and model reuse: a polemic. In E. Yiicesan, C.H. Chen, J.L. Snowdon & J. M. Charnes (Eds.), *Proceedings of the 2002 Winter Simulation Conference*: Institude of electrical engineers.

Pidd, M. (2004). *Computer Simulation in Management Science*. London: John Wiley and Sons.

Robinson, S. (1994). *Successful Simulation*: John Wiley and Sons.

Robinson, S. (2004). *Simulation: The Practice of Model Development and Use*: John Wiley and Sons.

Robinson, S., Nance, R., E., Paul, R. J., Pidd, M., & Taylor, S. J. E. (2004). Simulation model reuse: definitions, benefits and obstacles. *Simulation Modelling Practice and Theory, 12*, 479-494.

Robinson, S., Radnor, Z. J., Burgess, N., & Worthington, C. (2012). SimLean: Utilising simulation in the implementation of lean in healthcare. *European Journal of Operational Research, 219*, 188-197.

Rouwette, E. A. J. A., Größler, A., & Vennix, J. A. M. (2004). Exploring Influencing Factors on Rationality: A Literature Review of Dynamic Decision-Making Studies in System Dynamics. *Systems Research and Behavioral Science, 21*, 351-370.

Rouwette, E. A. J. A., Korzilius, H., Vennix, J., & Jacobs, E. (2011). Modeling as Persuasion: the impact of group model building on attitudes and behaviour. *System Dynamics Review, 27*, 1-21.

Rouwette, E. A. J. A., Vennix, J. A. M., & Mullemkom, T. (2002). Group model building effectiveness: a review of assessment studies. *System Dynamics Review, 18*, 5.

Simul8. (2009). Simul8 Software. In.

Sterman, J. D. (1989). Misperceptions of Feedback in Dynamic Decision Making. *Organizational Behavior and Human Decision Processes*, 301-335.

Su, X. (2008). Bounded rationality in newsvendor models. *Manufacturing and Service Operations Management, 10*, 566-589.

Suri, R. (1998). *Quick Response Manufacturing. A Companywide approach to reducing lead times*. Portland, Oregon: Productivity Press.

Tako, A. A., & Robinson, S. (2010). Model development in discrete-event simulation and system dynamics: An empirical study of expert modellers. *European Journal of Operational Research, 207*, 784-794.

Thomke, S., H. (1998). Simulation, Learning and R&D performance: Evidence from automotive development. *Research Policy, 27*, 55-74.

Tulving, E., & Kroll, N. (1995). Novelty assessment in the brain and long-term memory encoding. *Psychonomic Bulletin and Review, 2*, 387-390.

Tversky, A., & Kahneman , D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Van der Zee, D.-J., & Slomp, J. (2009). Simulation as a tool for gaming and training in operations management - a case study. *Journal of Simulation, 3*, 17-28.

Ward, S. C. (1989). Arguments for Constructively Simple Models. *The Journal of the Operational Research Soceity*, 141-153.

Yaniv, L. (2004). Recieving other people's advice: Influence and Benefit. *Organisational Science and Human Decision Processes, 93*, 1-13.