Proceedings of the Operational Research Society Simulation Workshop 2018 (SW18)
A.Anagnostou, R.Meskarian, D.Robertson eds.

CAN WE LEARN FROM WRONG SIMULATION MODELS? A PRELIMINARY EXPERIMENTAL STUDY ON USER LEARNING

Mr. Naoum Tsioptsias

School of Business & Economics
Loughborough University
Epinal Way, Loughborough LE11 3TU
n.tsioptsias@lboro.ac.uk

Dr. Antuela Tako

School of Business & Economics
Loughborough University
Epinal Way, Loughborough LE11 3TU
a.takou@lboro.ac.uk

Professor Stewart Robinson

School of Business & Economics
Loughborough University
Epinal Way, Loughborough LE11 3TU
s.l.robinson@lboro.ac.uk

ABSTRACT

A number of authors believe that wrong models can be useful, providing learning opportunities for their users. This paper details an experiment on model complexity, investigating differences in learning after using a simplified versus an adequate version of the same model. Undergraduate students were asked to solve a resource utilization task for an ambulance service. The treatment variables were defined as the model types used (complex, simple, and no model). Two questionnaires (before and after the process) and a presentation captured participants' attitudes towards the solution. Results suggest differences in learning were not significant, while simple model users demonstrated a better understanding of the problem. This paper consists of a preliminary behavioural operational research study that contributes towards identifying the value of wrong simulation models from the perspective of model users.

Keywords: Discrete-Event Simulation, wrong models, Simple models, Complexity, Behavioural Operational Research

1 INTRODUCTION

Models are simplified representations of real life situations, which one could call simply "wrong" (Box and Draper, 1987). However, models have been used to support decision-making and problem-solving in organisational activities, both in business and governmental areas (Luoma, 2014; Pace, 2004; Wahlström, 1994). Understanding model use from the clients' perspective in simulation was first studied in Tako and Robinson (2009). Since then, with the emergence of Behavioural Operational Research (BOR), more interest has risen in understanding model acceptance from the clients' point of view (e.g. Hämäläinen et al., 2013; Gogi et al., 2016; Monks et al., 2016). More recently, Katsikopoulos et al. (2017) consider the benefits from using simple versus complex models in decision making, albeit from the point of view of multi-criteria decisions analysis. Using simplified models can affect clients' perception of model validity and the model may be considered inadequate or plainly wrong to use. Yet, even if a model is considered wrong, it is believed that we can still learn from it (Hodges, 1991; Bankes, 1998).

So far we have not found explicit definitions of wrong models in the existing literature. To the best of our knowledge, there is no evidence to demonstrate the benefits, if there are any, of using simulation models considered wrong on users' learning. This paper explores the concept of "wrong" models, and more specifically, we concentrate on model complexity from the client's perspective. We look to establish whether the level of model complexity affects the learning achieved by users of discrete event simulation models. This will in turn provide evidence to inform our understanding of model use in supporting learning from using simple versus complex models. This can be especially relevant for the existing facilitated modelling practice (e.g. Franco and Montibeller, 2010; Tako et al., 2010), where it may not be possible to revisit and rebuild a model at the workshop.

We present a preliminary experimental study carried out with undergraduate students at Loughborough University to identify differences in participants' learning as a result of using a simple versus a more complex version of the same model. Our study aims to provide evidence on the usefulness of wrong - or as termed here, "simple" - models. It is part of a wider study looking at the uses of wrong models and their role in supporting learning and decision-making. This work contributes to the existing BOR and simulation literature towards identifying the value of wrong models from the perspective of model users.

This paper is organised as follows: Section 2 summarises the existing work on wrong models in Operational Research (OR), Simulation and relevant fields, also introducing learning. Section 3 presents the methodology, aims and hypothesis, the case study and the process followed, while Section 4 reports the results. Section 5 provides a discussion of the findings followed by the conclusion in Section 6.

2 WRONG MODELS IN OR AND SIMULATION

Literature in OR and simulation is limited on the topic of wrong models, mainly focusing on successful cases (Bankes, 1993; Eskinazi and Fokkema, 2006). There are limited examples of wrongly developed models, from which we could learn. An exception is the work by Eskinazi and Fokkema (2006), where the authors demonstrate four failed modelling interventions in System Dynamics (SD), though their qualitative analysis doesn't reach any conclusive outcomes.

Considering the terminology of wrong models in the existing literature, our first observation is that there is no clear definition. A variety of terms are used, among others: bad (Hodges, 1991), unvalidatable (Hodges and Dewar, 1992), inadequate or unvalidated (Hodges, 1991), incorrect (Bankes, 1998), false (Bankes, 1998; Hooker, 2007) or wrong (Hinkkanen et al., 1995; Bankes, 1998). More specifically, Hodges (1991) was the first to refer to the concept of wrong models in policy analysis. With reference to military examples, Hodges (1991) distinguishes wrong models as "unvalidated" (not adequately validated for their intended purpose) or "invalidated" (having failed validation). In a subsequent paper, the term "unvalidatable" is used for models that cannot be validated but may still be utilised (Hodges and Dewar, 1992). A categorization of models is given by Bankes (1998) terming them as false ("demonstrably incorrect"), strongly predictive, and plausible or weakly predictive (models wrong in at least some aspect). Hooker (2007) suggests that a false model may not predict the system it was created for. Sasou et al. (1996) examine the term from the point of view of a team's wrong decision-making process.

Examining the factors that may lead to a wrong model, Hodges (1991) refers to problems with data such as contradictions, lack of data or inconsistencies as determinants that could lead to bad models. Bankes (1993) mentions that it may not be possible to validate because experiments are not feasible to be carried out, historical data may not be available, or there is no sufficiently developed theory to support model assumptions. The same author (Bankes, 1998) refers to conflicts deriving from internal model structure and behaviour that contradict existing knowledge. Robinson (1999) mentions causes of simulation inaccuracies created by insufficient experimentation. From a different viewpoint, a model may be wrong because it is used for solving the wrong problem (Balci, 1994; Hooker, 2007).

Furthermore the literature refers to uses of wrong models. There is some consensus among authors that wrong models can still be useful and can be used creatively (Hodges, 1991; Bankes, 1993; Bankes, 1998; Eskinazi and Fokkema, 2006). Hodges (1991) and Hodges and Dewar (1992) suggest a number of possible implementations without further practical applications. Similarly, Bankes (1993; 1998) discusses that exploratory (not restricted by uncertainty issues) and plausible (weakly predictive) models can still assist in decision making.

The use of a model - either simple or complex - provides learning opportunities for its users. Learning produces a change in behaviour by deflecting someone's observable action (Schacter et al., 2014). Based on Argyris and Schön (1996), one can deduce that learning is achieved if a change in users' existing knowledge, attitude or decisions occurs as a result of interacting with a model. Recent studies have initiated the practical exploration of learning in the existing BOR field. Monks et al. (2014) test learning differences between reusing a model versus involving clients in model building. Monks et al., (2016) expand this premise by comparing learning transfer between simulation studies. Gogi et al. (2016) search for learning insights from using discrete event simulation. Though not conclusive, all three studies contain positive findings that corroborate the need for further investigation of learning using simulation models.

In summary, based on the work considered above, the uses of wrong models have not yet been clearly identified in the literature. Here we consider one element of wrong models, the level of model complexity from the clients' perspective, distinguished into simple and complex models. There is no evidence to suggest what level of model complexity should be aimed to achieve the best possible learning outcomes. Studies in OR mainly in the field of forecasting, explore the idea of simple versus complex but in view of better results (e.g. Green and Armstrong, 2015). Katsikopoulos et al. (2017) summarise relevant studies (in inference, forecasting and strategic decision-making) and provide guidelines to help decide when simple models should be used for specific decisions, concluding that simple models should be used at the right level of simplification. Our study is looking to compare learning achieved from the users' point of view after using models of the same problem at different levels of complexity. Hence the current study looks to explore the differences in learning from using simple versus complex simulation models.

3 METHODOLOGY AND EXPERIMENTAL DESIGN

This section describes the experimental study, including the research questions and hypothesis, study design, the case study, and the simulation models used.

3.1 Study objective and hypothesis

The aim of this research is to identify whether a model's level of complexity affects the learning achieved by model users. We consider complexity from the point of view that less complexity can affect clients' perception of the model, since increasing a model's complexity doesn't necessarily mean that the model is more accurate (Robinson, 2014). We link this to the notion of model credibility, that is the clients' perception that the model or its results are sufficiently accurate for the purpose at hand (Pace, 2004; Robinson, 2014). As a result through validation procedures, modellers try to prove that the model is not perceived wrong (Robinson, 2014).

We consider as wrong a simplified model (also called simple) and compare it with an adequate (complex) model. In this research, we assume that learning occurs as a result of a change in people's attitude towards a belief. This can be demonstrated by providing the expected answer to a problem. As such we look to test the following hypothesis:

Study hypothesis: The use of simple and complex simulation models offers the same learning outcome.

We expect that users will find simple models easier to use and hence provide evidence that they gain better understanding of the problem and solutions, based on statements found in the literature (e.g. Green and Armstrong, 2014; Katsikopoulos et al., 2017).

3.2 Study design

An experiment is used to test the study hypothesis. Final year undergraduate students, 58 in total, at Loughborough University attending a simulation module "Simulation for Decision Support", took part in the experiment. The experiment was run as part of a three-hour lecture. All students were aware of basic simulation modelling and had undertaken a placement year in a company in the third year of their studies. They were separated in groups of 6 or 7, for a total of 9 groups. Each group was assigned to one of the treatment conditions. These conditions were defined by the type of model the participants used: Complex Model (CM) groups used a relatively complete or adequate simulation model, Simple Model (SM) groups a simplified (wrong) simulation model and No Model (NM) groups, consisting the control groups, were asked to create a conceptual model of the case study problem. The participants were not aware of any model or condition differences. Table 1 summarises the assignment of group numbers into treatment conditions:

	Condition	Abbreviation	Groups assigned	Model provided
1	Complex Model	CM	2, 3, 4	Adequate model
2	Simple Model	SM	1, 5, 6	Simplified simulation model - (can be potentially perceived wrong)
3	No Model	NM	7, 8, 9	No model

Table 1 *Group assignment to treatment conditions*

The process took place in the following order. An overview of the problem and tasks was provided. The participants were then given the case study to read along with the pre-test questionnaire to complete individually, capturing participants' initial attitude towards the problem and the managerial decision. Next, they were randomly split into groups. Groups 1 to 6 were each provided with a notebook computer with the allocated version of the model (simple or complex). Further paper-based instructions were given to each group based on the treatment condition they were assigned to. The students worked in groups and were asked to utilise PartiSim tools (Tako et al., 2010) to guide their group' discussions. The students were left to work on their task for 1.5hrs and to prepare a presentation with their recommendations towards the case scenarios (more details are provided in Section 3.3). During the task, support was provided by the researchers with clarifications on the process and task. At the end the students re-assembled in the lecture theatre and were asked to complete an individual post-test questionnaire, in order to capture attitude changes towards the solution. The questionnaire was the same for all groups, with the exception that the No Model groups were asked not to answer two questions related to the model. Lastly, the students presented their findings to the other groups and the researchers. Questions were asked by the researchers to clarify the students' final answer towards the decision and the prioritisation of targets.

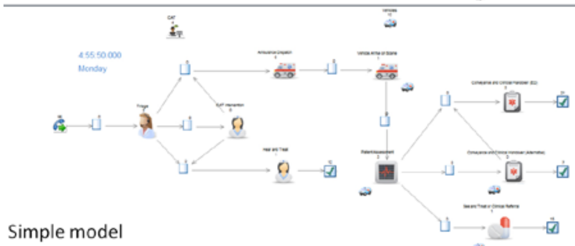
3.3 Case study

The case study is a resource utilisation task for an ambulance service based on Puntambekar (2016). The case initially describes the problem and the call cycle. Incoming calls are classified as emergencies (life-threatening) or urgent (non-life-threatening). When answering a call, the operators assess the severity of patients' condition and decide on the route to be followed. Regardless of call type, a proportion of calls is redirected to the Clinical Assessment Team (CAT) for re-evaluation. The majority of calls result in the patients being transported to the local A&E department or to alternative pathways (e.g. community care services). In some cases, the ambulance crew provides clinical treatment on scene and the patient may not need to be conveyed to A&E. Treatment may be also provided over the phone by clinical advisors. This helps to avoid the dispatch of ambulances to patients who do not require an ambulance. During the winter months, the ambulance service faces a higher number of calls, which affects the service's ability to deal with incoming calls within specified time targets. The management considers that a lot of patients are unnecessarily being taken to A&E and are examining the option to increase CAT intervention with the view to reducing ambulance use, and open up resources for patients that require emergency transportation. Three options are available: keeping the percentage as is (30%), increasing it slightly (40%) or significantly (50%). Either increase was described as possible. The task asked participants to comment on this managerial decision and recommend what percentage of patient calls should be redirected for re-evaluation by the CAT team. To check which answer would be the most suitable, certain targets were set that the participants would have to meet and prioritise. These targets (in order of importance) were: the time targets for life-threatening patients, the time target for non-life-threatening patients, the maximum expenditures (costs of personnel, ambulances), and, the ratio of the number of patients treated in A&E over those treated in alternative pathways. The full case study is available by the authors upon request.

3.4 The two variations of the model

The main model used was created as part of an MSc dissertation project in Business Analytics and Consulting through facilitated workshops with employees of the ambulance service (Puntambekar, 2016). The model was modified and financial variables were added. The Simul8 software (SIMUL8 Corporation) was used to develop the models. Two main variations (Figure 2) of that same model were selected: an adequate termed here as "complex" and a simplified one termed here as "simple". The simple model was created from the originally developed model by taking out variables, parameters, working stations, and simplifying the routes (for example the transport of a patient to an alternative care would require a request in the complex model before proceeding, while the simple model omits this step). It was functional but less detailed. For instance, the types of available resources were reduced from three (CAT personnel, First Vehicles on Scene, and Last Vehicles on Scene) to two (CAT personnel, and Vehicles). As a result of these simplifications, the numerical outcomes were not as accurate as those in the adequate model. Both models could provide answers for the initial decision (30%) if participants were to change certain parameters. This means that the participants could reach all of the targets if they changed the variables of the simulation models.

Type of model	Settings	Solvable	Time required	Expected answer	Interpretation
Complex	30%	Yes	Low	Yes	Change of attitude
	40%	Yes	High	No	Luck or guessing
	50%	Yes	High	No	Luck or guessing



Type of model	Settings	Solvable	Time required	Expected answer	Interpretation
Simple	30%	Yes	Low	Yes	Change of attitude
	40%	Yes	High	No	Luck or guessing
	50%	No	N/A	No	Guessing

Figure 1 *The two variations of the model and the answers and interpretations per case*

Using the complex model, the problem could be solved for both alternatives (40% and 50%), but it required too long to trial out all possible options. The simple model could not provide answers to the problem for the 50% option (meaning that not all the targets could be met for that percentage - column "Solvable" in Figure 1). In order for the participants to demonstrate a change of attitude (leading to learning), they would need to understand - through experimentation with the model - that the initial setting (30%) is the answer that offers a solution to the model by meeting all of the targets without extreme effort (columns "Time required" & "Expected answer" in Figure 2). Since the alternative settings (40% and 50%) would require too much time to run, if such an answer was provided and not supported by meeting the targets, it would be interpreted as either guessing from the participants' side (column "Interpretation" in Figure 2) or luck if they managed to meet all of the targets within the limited timeframe. To test the experiment's feasibility, 3 pilots were run individually with PhD students. A few minor changes in the description of the case study were made as a result.

4 RESULTS

In total, 39 pre-test and 45 post-test questionnaires were returned from the 58 initial participants. The 6 additional post-test responses were excluded from the analysis. Of these, the NM group had 10 students, SM 16 and the CM the remaining 13 students. All students were 21-23 years old (but one who was 25). The rest of the demographics are provided in Table 1. The distribution of abilities and marks is representative amongst the different treatment groups.

	Groups	**Participants**	**Gender**		**Current mark of degree**		
			M	**F**	**1st**	**2:1**	**2:2**
1	No Model (NM)	10	15.4%	10.3%	7.7%	12.8%	5.1%
2	Simple Model (SM)	16	25.6%	15.4%	5.1%	28.2%	7.7%
3	Complex Model (CM)	13	28.2%	5.1%	7.7%	20.6%	5.1%
4	All	39	69.2%	30.8%	20.5%	61.6%	17.9%

Table 2 *Demographics of the participants that handed over both questionnaires. Proportions are calculated in every case out of the 39 participants*

The results are next analysed. To test the study hypothesis, we compare participants' answers based on the solutions provided before and after using the models. If there is no significant difference between the users of SM and CM, then the research hypothesis can be supported. We also compare the two simulation conditions with NM to establish whether there is a difference between using the model at all, as a means of checking that the case study and model work. If simulation users demonstrate a shift in their attitudes towards the solution as opposed to NM, then we can support that the case study and model work. We use Pearson's Chi-square and Fisher's Exact Test for comparing our nominal variables (Section 4.1), and the Mann-Whitney test for our likert-based variables (Section 4.2). All statistical tests are selected based on their relevance to the type of data analysed and the number of the groups compared ($\alpha = 5\%$) and are based on Bryman and Cramer (2011). First the results based on the managerial decision (Section 4.1) are presented, then the analysis the Likert scale questions (Section 4.2), and lastly the outcomes from the presentations (Section 4.3).

4.1 Results on the managerial decision

The participants were asked both in the pre- and post-questionnaire to answer whether and why they agree, disagree, or are not sure about the proposed managerial decision to increase the percentage of patients that are redirected to the CAT team for re-evaluation. The answers to this question are compared to establish a change in participants' perception of the solution. As expected the participants did not disagree with the intended managerial decision in the pre-test. Indeed, 24 were in agreement and 15 were not sure (0 disagreed), as opposed to 21 agreed, 10 disagreed, and 8 were not sure in the post-test.

For learning to be achieved, students had to move to a better decision by disagreeing or at least expressing uncertainty about the increase of this percentage. Our premise was that if students altered their initial views from "Agree" or "Not Sure" to "Disagree", or from "Agree" to "Not Sure" then they would have acquired a change in beliefs through the process. This change - observable through their answers - would support the presence of learning (Argyris and Schön, 1996; Schacter et al., 2014). It should be noted that if a participant was not sure of the answer both before and after the experiment, his belief was not considered to have changed. The participants' changes in beliefs, by group, are summarised in Table 3.

	Groups	Change	No change	Participants in groups
1	No model groups	0 (0%)	10 (25.6%)	10
2	Simple model groups	7 (17.9%)	9 (23.1%)	16
3	Complex model groups	8 (20.5%)	5 (12.8%)	13
4	Total	15 (38.5%)	24 (61.5%)	39

Table 3 *Change of participants' attitude based on comparison of pre- and post-test answers. Proportions are calculated in every case out of the 39 participants*

A shift in the participants' answers is noted. From the 39 participants, 15 of them (38.5%) changed their initial views towards the required solution. More specifically, 8 participants were from the CM groups, 7 from the SM groups, and no one (as expected) from the NM groups. This means that 15 out of 29 (51.7%) simulation group participants managed to find the required answer in view of the problem's managerial decision (i.e. keep the percentage at 30%). We used the Chi-square test to compare the proportion of participants that change beliefs about the managerial decision between groups. Due to the small sample size, we also report Fisher's exact test (Table 4). The results show that there is a difference in change of beliefs between the NM groups and those that used a simulation model (Pearson's Chi-square p = .004, Fisher's exact test p = .006), meaning that simulation had an impact in the experimental process. On the other hand, the results didn't reveal any actual differences between complex and simple model users (Pearson's Chi-square p = 0.340, Fisher's Exact Test p = 0.462) supporting the research hypothesis.

Group comparison	Pearson's Chi-Square	Fisher's Exact Test
NM vs SM & CM	0.004	0.006
NM vs SM	0.014	0.023
NM vs CM	0.002	0.003
CM vs SM	0.340	0.462

NM: No Model, SM: Simple Model, CM: Complex Model

Table 4 *Statistical analysis at 5% for the managerial decision*

The explanations on the answers provided by the participants about the proposed managerial decision to increase the percentage of patients that are redirected to the CAT for re-evaluation in the pre- and post-questionnaire, helped to establish a better understanding of the group differences. On the one hand, those agreeing with the managerial decision for increasing the percentage supported their views on the premise of the case study's description (faster, better, and cheaper system). It should be noted, however, that certain participants were "not sure" of the managerial decision in the pre-test questionnaire misdoubting the description (e.g. "change may not have an impact" or "not enough information in statistics or simulation"). On the contrary, participants that changed their attitude towards the managerial decision and disagreed or stated that they were not sure about the increase, mainly commented on the numeric outcomes of their experimentation, on logical statements (e.g. "we can reach all targets without increasing"), or due to the model's simplifications, a need for further statistics, or that the model may already work without changes. These answers advocate that the research hypothesis is supported as we had change of beliefs during the experiment as a result of the treatment conditions and these changes were supported with elaborated answers and statistical significance.

4.2 Analysis of Likert-style questions

The post-test questionnaire included a number of 5-point Likert scale questions where participants were asked to rate their level of confidence in the model as well as their opinion about their understanding of the model, model representativeness and trust in results. A 5-point Likert scale was used. We used the Mann-Whitney test to analyse the differences and to establish whether there is a difference between groups. The results are presented in Table 5 below. All the results suggest that neither confidence nor understanding was affected by the experimental conditions. Similarly, there was no statistically significant difference between the users that used a simulation model and those that didn't. Participants' confidence in their answers and belief in the results was not affected by the treatment conditions.

We compare the participants' opinions about the model's representativeness and their trust in model results only for the two simulation model treatment groups (groups 1-6) using the Mann-Whitney test. Considering participants' trust in model results, there is no difference between groups ($p = 0.105$). Participants' opinions about model representativeness show a marginal difference suggesting that the participants' perception of the model may have been affected by the two different treatment conditions ($p = 0.055$). This is relevant to comments found from the qualitative analysis (see Section 4.1) and the presentations where participants from the simple model groups suggested they needed a better model while 2 out of 3 groups said the models are not very representative but could work with them (see more in Section 4.3).

Group comparison	**Confidence (Mann-Whitney)**	**Understanding (Mann-Whitney)**	**Representativeness (Mann-Whitney)**	**Trust in Results (Mann-Whitney)**
NM vs SM & CM	0.600	0.216	N/A	N/A
NM vs SM	0.636	0.615	N/A	N/A
NM vs CM	0.647	0.080	N/A	N/A
CM vs SM	0.931	0.098	0.055	0.105

NM: No Model, SM: Simple Model, CM: Complex Model
No results were found to be statistically significant at $\alpha = 5\%$

Table 5 *Statistical analysis at 5% for Likert-style variables measured*

4.3 Group presentations

Participants gave groups presentations at the end of the session. An incentive was provided for the best two presentations and the performance was rated by two of the authors based on answers' insightfulness and general format of each presentation. The outcomes and main points of the presentations are presented in Table 6 below:

	No Model	**Complex Model**	**Simple Model**
Correct priorities	2 out of 3	3 out of 3	3 out of 3
Final answer	0 out of 3 right	1 out of 3 right	2 out of 3 right
Targets	2 out of 3 (representative conceptual model)	3 out of 3 (not all of targets solved)	3 out of 3 (not all of targets solved)
Model evaluation	N/A	N/A	2 out of 3 (needed a better model)

Table 6 *Summary of the presentations*

In general the presentations showed a good understanding of the problem and the task. All three NM groups worked on creating a conceptual model that would represent the case study, with 2 out of 3 creating a very representative one (compared to the complex simulation model), while 2 out of 3 prioritised the targets in the required order. The CM groups presented controversial results. Although the use of simulation helped their better understanding in the target order (with minor differences), their final answers were not based on solving all targets (2 out of 3 groups suggested that the managerial decision for increase should be put to action at the 40% level). This means that instead of trying to solve the initial problem through minor changes they redirected their attention to the alternatives which were way more difficult to be accomplished. Only one group had the correct final answer presented and justified. The SM groups got a well justified order of priorities (with minor differences). Still, no group met all the targets set, but 2 out of 3 groups replied that their proposal to the management would be to not change the current percentage for the managerial decision. This means that by using a simpler version of the model the participants got the right answer with more ease, though the most interesting outcome from these groups' presentations was that 2 out of 3 commented about the fact that the model was not representative as the reason that they were not able to provide complete answers or for not being able to meet the targets. A group members, specifically, reported that they felt a lot of information was missing from the model they were given (not knowing it was the simplified version). This is an interesting finding that suggests that the model given was considered wrong from a credibility point of view, it was however still adequate to use to find the relevant solutions to the problem.

5 DISCUSSION ON RESULTS

We present an experimental study set out to identify whether a model's level of complexity affects the learning achieved by users. The hypothesis of the experiment suggests that a simple simulation model compared to a more complex can offer the same level of learning. Though in small scale, this preliminary experiment allows the inductions of a few observations for further exploration of wrong models.

To support the hypothesis, the users of the simple and the complex model would demonstrate the same change of attitude towards the solution. The results show that a sufficient number of students from both treatment conditions provided the relevant solutions (8/13 complex model users and 7/16 simple model users). This means that both models proved useful to help participants understand the problem. The statistical analysis supports our hypothesis that there is no significant difference in the users' confidence in their answers, their belief in the results and their understanding of the problem. In the group presentations, one complex model group and two simple model groups managed to find the correct answer, while two of the simple model groups mentioned they found their model was lacking details but could still work with it. Our principal observation is that the simple model seemed easier for participants to handle especially due to the limited timeframe required to interact with the model. This outcome justifies our initial expectations. Furthermore, a marginally non-statistically significant result on representativeness (5.5%) suggests model users may have realised the adequacy of their model in their respective groups. The analysis corroborates our belief that the simplified model was not credible (users found it "wrong") but it still proves to be helpful. Another observation from our experiment regards the fact that simulation helped model users to demonstrate a shift in their attitudes towards the solutions compared to the control groups. Despite confidence and understanding were not found to be statistically different between the no model and simulation users groups, we still believe that simulation helped.

There are certain limitations related to the study that may have affected the findings. Not all students provided elaborate answers to the open-ended questions, leading to a small number of answers to analyse, which has affected the information collected. Due to the relatively small sample size we only tested one model at two complexity levels. If we had access to a larger group, we would have been able to test different levels of model complexity. A future addition to the experiment could be to include post-graduate and research students. Furthermore, group composition, dynamics and dysfunctionality may have affected the outcomes of group results and the quality of participants' answers. Bearing in mind these limitations, we next plan to run another set of experiments individually with participants in order to limit the impact of group-related factors. An extension to this study would be to identify the minimum requirements for a wrong model to be considered useful to model users for learning purposes.

6 CONCLUSIONS

This paper explores the concept of wrong models focusing on complexity. A laboratory experiment was set up to identify whether model complexity can affect the learning achieved. We compare two types of models: an adequate and a simplified one. The results suggest differences in learning from the two models were not significant. We however found that simple model users had a better understanding of the problem, albeit they were able to comment that they needed a more detailed model to be able to solve the task set. These initial results are encouraging, providing some preliminary evidence that simple models, so long as they are not inaccurately presented, can be useful in supporting clients to understand their problems and take decisions. This work can be particularly useful to inform the current facilitated modelling practice with regards to the type of models and complexity used in workshops with clients. Future research will aim to identify uses of wrong models by interviewing simulation experts.

REFERENCES

Argyris, C., and Schön, D. A. (1996). *Organizational Learning II - Theory, Method, and Practice*. Addison-Wesley.

Balci, O. (1994). Validation, verification, and testing techniques throughout the life cycle of a simulation study. *Annals of Operations Research*, 53, pp 121–173.

Bankes, S. (1993). Exploratory Modeling for Policy Analysis. *Operations Research*, *41*(3), pp 435–449.

Bankes, S. (1998). Policy Analysis for Complex and Uncertain Systems Through Computational Experiments. *Proceedings of IEEE Aerospace Conference*, pp. 9–21 vol.1.

Box, G. and Draper, N. (1987). *Empirical model-building and response surfaces*. New York: Wiley.

Bryman, A. and Cramer, D. (2011). *Quantitative Data Analysis with IBM SPSS 17, 18 & 19*. 1st edition. Hoboken: Routledge.

Eskinasi, M., and Fokkema, E. (2006). Lessons learned from unsuccessful modelling interventions. *Systems Research and Behavioral Science*, *23*, pp 483–492.

Franco, L. A., and Montibeller, G. (2010). Facilitated modelling in operational research. *European Journal of Operational Research*, *205*(3), pp 489–500.

Gogi, A., Tako, A. A., and Robinson, S. (2016). An experimental investigation into the role of simulation models in generating insights. *European Journal of Operational Research*, 249, pp 931–944.

Green, K. C., and Armstrong, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research*, *68*(8), pp 1678–1685.

Hämäläinen, R. P., Luoma, J., and Saarinen, E. (2013). On the importance of behavioral operational research: The case of understanding and communicating about dynamic systems. *European Journal of Operational Research*, 228(3), pp 623–634.

Hinkkanen, A., Lang, K. R., & Whinston, A. B. (1995). On the usage of qualitative reasoning as an approach towards enterprise modelling. *Annals of Operations Research*, *55*(1), pp 101–137.

Hodges, J., and Dewar, J. (1992). Is it you or your model talking? A framework for model validation. *http://www.rand.org/content/dam/rand/pubs/reports/2006/R4114.pdf*. Accessed 3 October 2015.

Hodges, J. S. (1991). Six (Or So) Things You Can Do with a Bad Model. *Operations Research*, 39(3), pp 355–365.

Hooker, J. N. (2007). Good and Bad Futures for Constraint Programming (and Operations Research). Constraint Programming Letters, 1, 21–32. *http://public.tepper.cmu.edu/jnh/cpfutures.pdf*. Accessed 10 October 2017.

Katsikopoulos, K. V., Durbach, I. N., and Stewart, T. J. (2017). When should we use simple decision models? A synthesis of various research strands. *Omega*, *0*, pp 1–9.

Luoma, J. (2014). Model-based organizational decision making: A behavioral lens. *European Journal of Operational Research*, *249*(3), pp 816–826.

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

Monks, T., Robinson, S., and Kotiadis, K. (2016). Can involving clients in simulation studies help them solve their future problems? A transfer of learning experiment. *European Journal of Operational Research*, 249(3), pp 919–930.

Pace, D. K. (2004). Modeling and simulation verification and validation challenges. *Johns Hopkins APL Technical Digest (Applied Physics Laboratory)*, *25*(2), pp 163–172.

Puntambekar, N. (2016). *Exploring the Efficiency of a UK Ambulance Service through Discrete-Event Simulation* (master's thesis). Loughborough University, UK.

Robinson, S. (1999). Three sources of simulation inaccuracy (and how to overcome them). *Proceedings of the 1999 Winter Simulation Conference*, pp 1701–1708.

Robinson, S. (2014). *Simulation - The Practice of Model Development and Use*, 2nd edition. Macmillan education, Palgrave.

Sasou, K., Takano, K., and Yoshimura, S. (1996). Modeling of a Team 's Decision-Making Process. *Safety Science*, *24*(1), pp 13–33.

Schacter, D., Gilbert, D., Wegner, D. and Hood, B. (2014). *Psychology*. 3rd ed. Worth Publishers.

Tako, A. A., Kotiadis, K., and Vasilakis, C. (2010). A Participative Modelling Framework for Developing Conceptual models in healthcare simulation studies. *Proceedings of the 2010 Winter Simulation conference*, pp. 500–512.

Tako, A. A., and Robinson, S. (2009). Comparing discrete-event simulation and system dynamics: users' perceptions. *Journal of the Operational Research Society*, *60*(3), pp 296–312.

Wahlström, B. (1994). Models, modelling and modellers: an application to risk analysis. *European Journal of Operational Research*, 75(3), pp 477–487.

AUTHOR BIOGRAPHIES

NAOUM TSIOPTSIAS is a Doctoral Researcher at Loughborough University, School of Business and Economics, in Operational Research and Simulation. He received a Bachelor on Finance & Accounting from the University of Macedonia (Thessaloniki) and an MsC (Hons) on Informatics & Management from the Aristotle University of Thessaloniki. His current research concerns the uses of wrong simulation models in OR, while previous interests regarded data mining on financial crisis. Before his PhD studies, he worked as an economist for an HR company.

ANTUELA TAKO is a Senior Lecturer in Operational Research at the School of Business and Economics, Loughborough University. She holds a PhD in Simulation and an MSc in Management Science and Operational Research from the University of Warwick. Her research interests include the comparison of simulation approaches (discrete-event simulation and system dynamics), facilitated and participative simulation modelling, conceptual modelling and health care modelling. She is an Associate Fellow of the Operational Research Society (AFORS).

STEWART ROBINSON is Dean and Professor of Management Science at Loughborough University, School of Business and Economics. Previously employed in simulation consultancy, he supported the use of simulation in companies throughout Europe and the rest of the world. He is author/co-author of six books on simulation. His research focuses on the practice of simulation model development and use. Key areas of interest are conceptual modelling, model validation, output analysis and alternative simulation methods (discrete-event, system dynamics and agent based). Professor Robinson is co-founder of the Journal of Simulation and the UK Simulation Workshop conference series. He was President of the Operational Research Society (2014-2015). Home page:www.stewartrobinson.co.uk.