

An Experimental Investigation into the Role of Simulation Models in Generating Insights

Anastasia Gogi, Antuela A. Tako and Stewart Robinson

Management Science and Operations Management Group
School of Business and Economics,
Loughborough University,
Loughborough,
Leicestershire,
LE11 3TU, UK

Corresponding Author

Email: s.l.robinson@lboro.ac.uk

Telephone: +44 (0)1509 222432

Abstract

It is often claimed that discrete-event simulation (DES) models are useful for generating insights. There is, however, almost no empirical evidence to support this claim. To address this issue we perform an experimental study which investigates the role of DES, specifically the simulation animation and statistical results, in generating insight (an 'Aha!' moment). Undergraduate students were placed in three separate groups and given a task to solve using a model with only animation, a model with only statistical results, or using no model at all. The task was based around the UK's NHS111 telephone service for non-emergency health care. Performance was measured based on whether participants solved the task with insight, the time taken to achieve insight and the participants' problem-solving patterns. The results show that there is some association between insight generation and the use of a simulation model, particularly the use of the statistical results generated from the model. While there is no evidence that insights were generated more frequently from statistical results than the use of animation, the participants using the statistical results generated insights more rapidly.

Keywords: Discrete-Event Simulation; Insight; Animation; Experimentation; Behavioural Operational Research

An Experimental Investigation into the Role of Simulation Models in Generating Insights

1 Introduction

Discrete-event simulation (DES) is a popular modelling technique that is claimed to support problem solving and decision making. Indeed, it is often said that clients gain ‘insights’ as a result of simulation interventions, especially from the simulation animation (Hurrion 1986; Belton and Elder 1994; de Vreede and Verbaeck 1996; O’Kane 2004; Proudlove, Black, and Fletcher 2007; van der Zee and Slomp 2009; Pidd 2010; Bayer et al. 2014). However, the term ‘insight’ is used quite loosely to mean an improved understanding. Cognitive psychologists explain that *insights* may refer not just to the acquisition of better understanding, but also to the experience of sudden shifts in understanding or ‘Aha!’ moments. More specifically, *insight* is defined as ‘*the cognitive process by which a problem solver suddenly moves from a state of not knowing how to solve a problem to a state of knowing how to solve that problem*’ (Mayer 2010, 276).

Given the claims about simulation in insight generation and the huge growth in simulation literature over the last two decades (Powers, Sanchez, and Lucas 2012), it is surprising that there is almost no empirical evidence to support the claims about insight. Evidence of learning outcomes is scarcely published in simulation papers (Fone et al. 2003). Even where the learning outcomes are reported, there is generally no explanation of the causal mechanism for learning, let alone Aha! moments. Therefore, any claim that the catalyst for insight is a simulation model, and more specifically the animated display of the simulation model, has relied largely on supposition and anecdotal evidence from case studies. Meanwhile, relatively little task-based behavioural research has been conducted aiming to support the above claims; and where it has, the results are mixed (Bell and O’Keefe 1995).

To address this dearth of evidence, this paper describes an experimental study that aims to test whether and how insights are generated from DES models. Our contribution is to provide a more in-depth understanding of insight in the context of simulation and empirical evidence on the role of simulation in generating insight.

The paper is organised as follows. In section 2, we describe in more detail the concept of insight and discuss how it relates to the simulation context. Then, we review the limited evidence that exists surrounding the use of simulation models and insight. In section 3, we present the experimental study, explaining the research hypotheses, the experimental design, the participants, the procedure, the dependent measures and the materials used. Section 4 details the results of the study, followed

by a discussion on the value of simulation models in insight generation, the limitations of the study and suggestions for future work (section 5).

2 Insight and Simulation

This section provides the conceptual foundation for our experimental study. We first introduce the concept of insight from relevant fields and we discuss its relevance to the simulation context. We then discuss the evidence that currently exists in the academic literature regarding the role of simulation in generating insight.

2.1 The Concept of Insight

The word 'insight' is used in two ways. It is used as a state of understanding – that is, to have insight into something (Smith 1995). Insight is also described as an experience, an Aha! experience, involving a moment of epiphany (Schooler, Fallshore, and Fiore 1995). This view is originally encountered in the story of Archimedes of Syracuse when he discovered the principle of displacement – 'eureka'. For this research we adopt this latter concept, proposing it as an approach to measure the value of simulation as a means for creating knowledge.

To explore Aha! insight in more depth, relevant literature is considered and in particular the theoretical domain of Gestalt theory (Maier 1940, Mayer 2010), creative cognitive psychology (Sternberg 2009) and a collection of studies on insight that have attempted to conceptualise the phenomenon (Metcalf and Wiebe 1987, Kounios and Beeman 2009). Despite the fact that these streams of literature do not share the same theoretical foundations, it seems that they all agree upon the phenomenological perspective of the concept: a satisfactory solution to a problem suddenly emerges after overcoming an impasse. An impasse is the state in which a problem solver realises that initial ideas do not solve the problem, but at the same time feels that all the possibilities have been exhausted (Schooler, Fallshore and Fiore 1995). Generating insight is a productive activity which is about doing something new or novel. New ideas that do not lead to the solution itself but are relevant to finding the solution are described as false insights (Isaak and Just 1995). They usually occur when the cause of a problem is misunderstood. In particular, in false insight, a person approaches the problem in a new or novel way, but without having a correct view of the problem. When false insights emerge, the suggested idea is not a satisfactory solution to the problem.

Insight differs from other problem solving approaches, such as intuition, which are often used synonymously in everyday speech. Dane and Pratt (2007) explain that while the concept of insight involves some degree of non-conscious thought, it arises through logical connections between a

problem and the solution. Intuition, in contrast, relies on non-conscious associative connections. Insight also differs from guessing in that the latter does not require making any sort of connection (i.e. conscious or unconscious). As a result, in problem solving with insight, a person is able to justify the suggested solution, whereas in problem solving with intuition, or guessing, the person is not.

Social scientists have offered many explanations about the mental mechanisms of insight generation which seem somewhat interrelated. In short, in achieving illumination (i.e. generating insight), a problem solver may overcome implicitly imposed constraints (Weisberg 1995), change mental representations, become aware of a new association between parts of a system, change the meaning of some problem element, or assimilate possible solutions from the environment (Davidson 1995). In other words, it is believed that prior knowledge and experience constrain people's worldview, and, as a result, this knowledge may prevent them from seeing the world as it really is. Nevertheless, by using past experience as a building block, avoiding being confined by habits or irrelevant associations, the problem solver may eventually identify the appropriate way to solve a problem; and hence insights emerge. For a more detailed discussion on insight see Gogi, Tako and Robinson (2014).

2.2 Insight in the Simulation Context

Applying the above in the context of DES, it can be claimed that insight occurs when people using a simulation model suddenly know how to improve the performance of a system after several failed alternative attempts (i.e. what-ifs scenarios). The strategy used to achieve major improvements in the system involves doing something new or novel. So, we can identify that insight occurs during the experimentation phase of a simulation project if the users go through a problem solving pattern where an impasse (i.e. a phase in which simulation users only run what-if scenarios which are similar to existing unsuccessful strategies) is followed by a sudden realisation/generation of new or novel ideas that achieve major improvements in the performance of as system. After the experience of an Aha! moment, the users must be able to justify the rationale underlying the suggested solution that has arisen. We note that since a simulation model is a simplification of the real world (Robinson 2008) insights only relate to the model, and they must be interpreted in the light of wider knowledge about the real-life system.

Table 1 provides a non-exhaustive list of instances of insight found in the simulation literature. For each example, a possible explanation about the mental mechanisms of experiencing insights is suggested with respect to the cognitive psychology theories introduced in section 2.1.

Table 1. Some examples of insights found in the simulation literature and possible mental mechanisms of experiencing insight.

Examples of Insight in the simulation literature	Possible Explanations of the Mental Mechanism of Experiencing Insight
<p>Robinson (2001)</p> <p>The stakeholders realise that there is no actual need for hiring lower skilled staff to work on the helpline</p>	<p><i>Overcoming implicitly imposed constraints</i></p>
<p>Lee (2010)</p> <p>Two subjects succeed in making significant improvements on their second representation of the water cycle compared to their initial one</p>	<p><i>Changing mental representation</i></p>
<p>Lee (2010)</p> <p>One subject builds new structural associations during the second representation of the water cycle</p>	<p><i>Becoming aware of a new association between parts of a system</i></p>
<p>Monks, Robinson and Kotiadis (2014)</p> <p>Some participants succeed in ceasing a common misconception about the relationship between resource utilisation and service level</p>	<p><i>Changing the meaning of some problem element</i></p>
<p>Bakken, Gould and Kim (1992)</p> <p>Some participants manage to apply insights learned from the first game to a second one</p>	<p><i>Assimilating possible solutions from the environment</i></p>

In this research, we use a single problem (i.e. the NHS111 case study as described in section 3.6.1) to explore the value of simulation in generating insights. As such, it is not possible to study assimilating possible solutions from the environment (i.e. there is no second game and consequently transferring knowledge acquired from one simulation game to another is not possible). In our case study, we also make explicit the parts of the system and all the associations between them. Therefore, it is not possible to study changing mental representation or becoming aware of some problem elements. Instead, we consider both implicitly imposed constraints and misconceptions of some problem element in order to study the value of simulation in generating insights.

2.3 Empirical Evidence on Insights with Simulation Models

There is very limited empirical evidence on the role of DES in learning and generating insight. This is despite Richels' (1981) suggestion, over thirty years ago, that models should be used as a means of education and exploration so that insights can be made available to clients. Indeed, van der Zee and Slomp (2009, 17) are surprised that simulation continues to be seen "as just a methodology to

analyse design decisions". Although Robinson (2005) warns that alternative uses of DES are crucial for the future of the technique, it is hard to find cases in the literature in which simulation deviates from its analytical problem-solving role. It seems that DES continues to be seen as just a 'hard' OR technique.

Practitioners and scholars often argue that DES interventions help stakeholders to gain insights about their problems and subsequently to generate effective ideas on how to address them (Hurrion 1986; Belton and Elder 1994; de Vreede and Verbaeck 1996; O'Kane 2004; Proudlove, Black, and Fletcher 2007; van der Zee and Slomp 2009; Pidd 2010; Bayer et al. 2014). Simulation case studies, however, rarely report on insights and even less on the stimuli that help people in real simulation settings arrive at insights. For instance, Taylor et al. (2009) perform a literature review of recent published simulation research and observe that there is a relative lack of research with real world involvement, and there is an even greater lack of evidence of the value of simulation in decision making. Other domain-specific reviews of the simulation literature have reached the same conclusion (Jun, Jacobsen and Swisher 1999; Fone et al. 2003; Günal and Pidd 2010).

Among the few examples of case studies that provide evidence of insight generation during a DES study are Elder (1992), Lehaney, Clarke and Paul (1999), Robinson (2001), den Hengst, de Vreede and Maghnouji (2007) and Fletcher et al. (2007). All these present real simulation projects in which the interest was not to use simulation as an analytical tool, but as a means to facilitate debate. In particular, the authors discuss how some model results, that were not anticipated, led to reassessment of initial plans and hence the generation of new and effective ideas on how to address a problem (i.e. insights).

While these studies provide some support that insights were generated during the interventions, it is difficult to identify the causal mechanisms of experiencing these Aha! moments. Because insight involves some degree of non-conscious thought, people appear unable to explain what elements of the DES intervention help them generate new ideas (Schooler, Fallshore, and Fiore 1995). Proponents of visual interactive simulation (VIS) (Crookes 1982; Bell and O'Keefe 1987; de Vreede and Verbaeck 1996; Paul, Giaglis, and Hlupic 1999; Macal 2001; Wenzel and Jessen 2001), which is the process of interacting with the graphical environment of a model during its execution for the purpose of model experimentation and analysis, argue that by watching the animation of a simulation model clients learn about their systems and so make better decisions. Again, the evidence to support this assertion is largely anecdotal and the empirical evidence is weak.

Laboratory based experimental studies could provide an empirical basis for exploring the claims about DES and insight. To our knowledge, however, the only elements of a DES intervention that have been tested in an experimental setting are the involvement of the client in the modelling process (Monks, Robinson, and Kotiadis 2014), the visual representation of a model (Carpenter et al. 1993; Swider, Bauer, and Schuppe 1994), and experimentation with a model (Luehrmann and Byrkett 1989, Parker 1991, O'Keefe and Pitt 1991, Chau and Bell 1994, Bell and O'Keefe 1995). These experimental studies provide useful findings about the role of simulation models in problem understanding and decision making. What they do not provide is a study on the generation of insight. Further, DES has developed significantly since the original studies were performed. For instance, in the work of Parker (1991) the simulation model did not include any animated displays (moving icons).

The experimental studies of Carpenter et al. (1993) and Swider, Bauer and Schuppe (1994) focus on the role of animation in facilitating communication with clients. The authors use students to determine which combinations of visual presentations (i.e. the movement, colour and detail of icons) and speed are best for communicating invalid model behaviour. The results suggest that animation with movement is better than animation with no movement or dynamically changing bar charts in communicating the operations of the simulation model. In addition, when the presentation speed is slow, response times are shorter with greater accuracy in problem identification.

In the experimental work of Bell and O'Keefe (1995), which is a refined version of O'Keefe and Pitt (1991), participants could choose among three different types of displays to solve a task: an animation (moving icons), a dynamically changing graphic and a numerical display (animation switched off). Although participants showed a strong preference for using the animation, its use did not result in better solutions and overall performance was poor. Nonetheless, participants who preferred to use the animated display tended to solve the problem with less effort (i.e. fewer alternatives) and quicker than participants who had switched off the animation.

Chau and Bell (1994) perform an experiment that aims to determine whether the use of VIS supports decision making. The authors compare task performance of students who made use of a VIS model to students who used a no-VIS model (animation is switched off and only numerical output are produced). The results reveal that the use of VIS is more effective than the traditional simulation model in terms of problem understanding, solution rates, solution time and number of scenario runs. However, in a similar experiment, Luehrmann and Byrkett (1989) find the opposite; participants who made use of a VIS model were slower and less accurate in their decisions than

participants who used a no-VIS model. The difference may be because Luehrmann and Byrnett only performed a small pilot study.

The most recent experimental study by Monks, Robinson and Kotiadis (2014) focuses on learning from DES studies and hints at a mechanism for generating insight. Through a laboratory experiment they specifically study the impact of model reuse versus model building on the client's ability to solve a task. They find that the clients appear to learn more from the use of the model than from their involvement in model building. However, new and effective ideas on how to address the problem were generated more frequently by participants who were involved in the model building. Although not identified by Monks, Robinson and Kotiadis, this suggests that involvement in model building is a mechanism for generating insight.

One particular stream of empirical research that is relevant to insight generation has been carried out by researchers in the system dynamics (SD) field. Researchers in this area are concerned with understanding how people learn from SD. Richardson et al (1994) operationalise learning as a mix of mental processes by which people change their mental models. The concept of mental model has been central to SD since the beginning of the field (Forrester 1975). Put simply, a mental model depicts a person's (or a group's) ideas on the structure of a system and how that is responsible for the system's behaviour. Behavioural decision theory reveals that people suffer from bounded rationality because of a number of judgemental biases and heuristics that people employ in complex situations (Kahneman 2011). Accordingly, people tend to misperceive dynamic rules and simplify the way a complex system is structured. That means that people's mental models are often unable to predict the behaviour of a dynamic system (Sterman 1994). Nonetheless, if people manage to refine their initial deficient mental models, learning occurs (Rouwette, Vennix and Felling 2004). As their mental models radically change, they may create new strategies to improve the performance of a system. Such productive activity denotes the use of the double-loop learning system as described by Argyris and Schön (1996) and is closely related to the concept of insight.

The counterintuitive behaviour of a system can mislead people about the actual causes of a problem. As a result they may not realise that their current strategies worsen the performance of the system (Forrester 1975). The analysis of a dynamic system with the use of a SD model provides a practical way to make explicit and test people's mental model (Hsaio and Richardson 1999). However, findings from experimental studies provide little evidence that the experimentation with a SD model support users in overcoming flaws in their mental model (Sterman 1989; Sterman 1994; Langley and Morecroft 2004; Monxes 2004; Rouwette, Größler and Vennix 2004; Sterman and Booth Sweeney 2007; Cronin, Gonzalez, and Sterman 2009). The studies show that participants often

appear unable to explain or intuitively predict the behaviour even of the simplest dynamic system. Rouwette et al (2011) provide some empirical evidence that SD interventions support people in changing their mental models and hence generating insights. However, it is not made explicit what elements of the intervention help produce those insights. Through an experimental study, Howie et al (2000) suggest that misperception of feedback is in part due to inferior interface design rather than just cognitive biases. The empirical evidence of this study shows that an improved human-computer interface can reduce the difficulties people have in dealing with complex systems, but it cannot eliminate them (i.e. optimal performance was not achieved by any participant).

Overall, the conclusion that can be drawn from this discussion is that although it is assumed that simulation models support problem solving and decision making, there is lack of empirical evidence on the efficacy of simulation models in generating insight. Similarly, the elements of a simulation intervention that stimulate insight generation are even less well understood. A specific claim is that watching the animated display of a simulation model is more helpful in making better decisions than relying on the statistical outcomes generated from simulation runs; but again, there is very limited evidence to support this.

3 Research Hypotheses and Approach

This section describes the details of the experimental study, explaining the hypotheses that are tested and the experimental design, and giving details of the participants, the experimental procedure, the dependent measures and the materials used for the experiment, including the case study based on the NHS111 telephone service for non-emergency health care.

3.1 Hypotheses

The objective of our research is to test whether simulation models, and specifically DES models, support users in generating insights. Our aim is to establish whether insights are generated primarily by watching the animated display of a simulation model or by using the statistical outcomes generated from simulation runs.

In order to address the above research objectives two hypotheses are examined.

Hypothesis 1: Insights are generated more frequently when a simulation model is used compared to when it is not used.

This hypothesis follows from the claim in the simulation literature that simulation can help generate insights (Hurrion 1986; Belton and Elder 1994; de Vreede and Verbaeck 1996; O’Kane 2004;

Proudlove, Black, and Fletcher 2007; van der Zee and Slomp 2009; Pidd 2010; Bayer et al. 2014) and from case studies in which simulation has helped people in generating new and effective ideas (Elder 1992; Lehane, Clarke, and Paul 1999; Robinson 2001; den Hengst, de Vreede, and Maghnouji 2007; Fletcher et al. 2007). In order to test this hypothesis we compare the frequency of insights generated by the participants who used a simulation model against the participants who did not. The latter participants formed the control group of this research. They were allowed to create scenarios on paper and then analyse them using their problem solving skills (calculators were permitted).

Hypothesis 2: The contribution of animation to the process of insight generation differs from the contribution of statistical outcomes.

Based on the work of Chau and Bell (1994) and Bell and O’Keefe (1995), our expectation is that the animated display is more helpful in generating insights than the statistical outcomes generated from simulation runs. However, due to a lack of recent evidence to support the above claim, and the opposite findings in Luehrmann and Byrkett (1989), hypothesis 2 has been set with no specific direction for the prediction. In order to test the contribution of animation versus statistical outcomes, we compare the frequency of insights, the task duration and the time to insight generated by the participants in the animation versus the statistics condition.

3.2 Design

We conducted a controlled experiment to meet the objectives of this study. We randomly assigned undergraduate students to two experimental conditions, namely *animation* and *statistics*, and asked them to solve a task either by using the animation or only the statistical outcomes of a simulation model. We compared their task performance to a more experienced group of undergraduate students who did not use the simulation model to solve the same task. The latter group is called *control group*, in that the subjects did not receive treatment (i.e. no simulation model was provided). As such, this study uses a quasi-experimental non-equivalent group design. Although the design is not completely randomised, quasi-experiments can still provide plausible causal knowledge about the impact of experimental factors (Thyer 2012, Campbell and Stanley 1963).

Results from a small-scale pilot study suggested that the proportion of people who solved the task without the use of a simulation model was 15% (2 out of 13). The subjects were all PhD students with varied backgrounds who voluntarily participated in the pilot study. Four were familiar with simulation, including the two solvers, and six had prior experience with the context of the case study (NHS111 or a similar service); among them was one solver. The results of the pilot study also

indicated that the proportion of solvers increased to 54% (7 out of 13) when a simulation model is used. Therefore, we estimated that a scientifically important difference in the proportion of solvers between the experimental conditions group and the control group is 0.40. This effect size may be categorised as medium (Cohen 1988). In accordance with Fleiss, Levin and Paik (2003, 76) we estimated that the required sample size is at least 20 per group at a significance level $\alpha = 0.05$ and power 80%.

3.3 Participants

The study was conducted at two universities in the UK. The two experimental conditions (i.e. simulation with *animation* or *statistics*) included in total 47 undergraduate students who took business, but no simulation modules, at Loughborough University. Students were randomly assigned to each experimental condition. The control group (i.e. no simulation model) of 20 participants consisted of undergraduate students who took simulation modules at Warwick University. All participants heard about the possibility to participate in this study through emails and announcement in lectures. To encourage participation in our study, monetary incentives were used (£10) (Abeler and Nosenzo 2013). An additional small monetary reward (£5), linked to the achievement of the task's goal, was given to all participants that provided valid solutions (Bonner and Sprinkle 2002). For the experimental conditions, participants were invited to attend one parallel session held in February 2014 and for the control group, one session was run in April 2014.

Demographic characteristics of the participants are presented in Table 2. There were no statistically significant differences between the experimental conditions group and the control group in terms of gender and prior experience with the context of the case study (NHS111 or a similar service). The selective criterion differed between the experimental conditions group and control group (i.e. taken simulation modules), and the groups also differed in degree year and in the number of modules taken with a quantitative content. In particular, Warwick students (control group) were more advanced in their studies and had taken more quantitative modules than Loughborough students (experimental conditions group). As such, if simulation does not affect task performance, we expected the control group to perform better than the experimental conditions group.

Table 2. Participants' characteristics.

	'Animation' Condition <i>Loughborough</i> <i>n = 25</i>		'Statistics' Condition <i>Loughborough</i> <i>n = 22</i>		Control Group <i>Warwick</i> <i>n = 20</i>		Significant Difference (experimental conditions to control group)?
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	
Gender:							No ($p = 0.658$) ^a
Male	13	52%	13	59%	9	45%	
Female	12	48%	9	41%	11	55%	
Prior Experience:							No ($p = 0.157$) ^a
Yes	2	8%	4	18%	6	30%	
No	23	92%	18	82%	14	70%	
Degree Year:							Yes ^b
1	16	64%	18	82%	0	0%	
2	9	36%	4	18%	10	50%	
3	0	0%	0	0%	10	50%	
Quantitative Modules:							Yes ^b
≤ 3	14	56%	15	68%	0	0%	
4-6	11	44%	7	32%	0	0%	
≥ 7	0	0%	0	0%	20	100%	

^a Differences in gender and prior experience between the groups were tested with chi-square tests.

^b Participants were not randomly assigned to the control group. As a result, chi-square test or Fisher exact test are not applicable for 'Degree Year' and 'Quantitative Modules' due to violation of contingency table analysis assumptions (i.e. structural zeroes and interaction between categories of the row and column variables) (Agresti 1990). From a simple table observation, it becomes evident that the number of Quantitative Modules predicts group assignment perfectly since Quantitative Modules < 7 corresponds to the Experimental Conditions and Quantitative Modules ≥ 7 corresponds to the Control Group. Degree year predicts group assignment perfectly when Degree Year is equal to 1 and 3. Therefore, there are significant differences in degree year and quantitative modules between the groups.

3.4 Procedure

The procedure followed during the experiment for the experimental conditions group and the control group is illustrated in Figure 1. Each subject was invited to attend one session in a classroom with all the participants of the same group they were allocated to. The two experimental conditions were run in parallel (the same date and time) so that information sharing between groups was prevented. One researcher and one invigilator were assigned for each session. The former led the session, following a structured pre-set script, and the latter made sure that participants worked individually and in silence. Initially, the participants were asked to read the case study. Next, subjects were given information about the process of the session and how to load the simulation model, if they were assigned to a group that was using the model. After that, all subjects completed a pre-test questionnaire. The questionnaire asked participants if they had prior experience with NHS111 or a similar service. Additionally, in order to assess their initial beliefs about the problem, participants

were asked: “Having read the description of the problem, why do you think NHS111 is not achieving its targets (give up to 3 reasons)?” Reading and completing the questionnaire took about 15

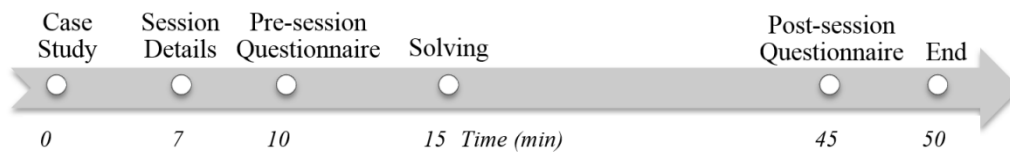


Figure 1. The procedure followed by participants in the experimental conditions and control groups on a time axis to indicate the starting point, duration and end point for each stage.

minutes.

After all participants had completed the questionnaire, they were provided with written instructions about how to approach the problem. The instructions differed slightly subject to the condition the participants were assigned to. In particular, participants in the animation condition were instructed to use the model by setting up scenarios and then watching the animated view of NHS111’s operations in the model. They could use three different speed levels according to their preference. At the end of a run, the performance (i.e. total cost and mean time in system) of each scenario was provided on screen which could be compared against the required targets (section 3.6.1). Participants in the statistics condition were similarly instructed to use the model as the participants in the animation condition, with the difference that the animation was turned off and they would only use the statistical results for each scenario such as time in the system, call waiting times, rates of abandoned calls and overestimated referrals. Participants in the control group were given a block of paper (‘scenario sheets’) and were instructed to create scenarios. They were asked to write down the reason they chose these scenarios and show some calculations (calculators were permitted) to confirm whether their scenarios solve the problem.

All participants were given 30 minutes to solve the problem. They were also informed that there is no limitation in the number of scenarios they could run. To ensure that individuals do not influence other participants’ thinking, they were instructed to ask questions in writing (an online platform was setup for participants in the experimental conditions group). When they finished, or when the time had expired, all participants were asked to submit their best scenario. Additionally, in order to assess their understanding at the end of the session, participants were asked the following two questions:

1. “Why do you think your suggested solution solves the NHS111 problem (give up to 3 reasons)?”
2. “What do you think NHS111 should do?”

3.5 Dependent Measures

For the experimental conditions group, solution rate, solution time and problem-solving patterns were derived from the computer recording the scenarios each participant attempted. For the control group, participants were asked to write down the scenarios and exact time that they began each scenario (a digital clock was placed in the front of the room). Participants' understanding before and after attempting to solve the problem was measured using the open-ended questions discussed in the previous section.

3.6 Materials

Three sets of materials were provided for the experiments: the case study, the simulation model and the instruction sheets. All these are available as supplementary material with this paper (the model as video files). Each is now described.

3.6.1 The Case Study

The case study is based around a fictional local NHS111 service; a telephone service for non-emergency healthcare in the UK. In brief, the service is largely manned by operators who are supported by a few expert clinical advisors (skilled nurses and highly skilled doctors). Operators are low-wage staff compared to clinical advisors. In particular, the wages of each type of staff for an evening shift are: £80 for an operator, £150 for a nurse and £240 for a doctor. In addition, an operator on average spends more time to assess the urgency of a call (i.e. 10 minutes/call) than the clinical advisors (i.e. a nurse on average needs 6 minutes/call and a doctor spends 5 minutes/call). They also often overestimate the urgency of call, or need to transfer it to clinical advisors due to their medical inexperience. As a result, the service has high operating costs and service times (Griffiths, Williams, and Wood 2013). Participants were asked to determine the appropriate number of staff so that required targets were met: not exceeding an evening shift budget of £2,200 and an average service time limit of 12 minutes.

In relation to the possible explanations about the mental mechanisms of insight generation presented in section 2.1, we consider that in our case study insights emerge after a participant overcomes implicitly imposed constraints and changes his/her understanding of the cause of the problem. In particular, we purposely made inconspicuous the possibility of removing a type of staff from the service so that subjects were likely to adopt constraints that were never explicitly stated. We also inferred that the evening shift service is understaffed compared to the volume of calls received in the morning shift so that participants were likely to initially misconceive the true cause of the service problem. These implicit constraints tend to lead participants into searching for a solution in a narrow problem space (i.e. scenarios including all types of staff) without realising that they were

looking in the wrong direction (i.e. assigning extra low-wage staff). As a result, they were led into an impasse which kept them from solving the problem (Isaak and Just 1995). Only when subjects realise that the real problem is not the lack of personnel, but the operators' inefficiency, and that the self-imposed constraint of 'the service must consist of all types of staff' can be broken, can the impasse be overcome. This could lead them rapidly to the solution (i.e. a service with no operators, only enough clinical advisors) by searching in the newly discovered problem space for the solution. Thus, the challenge of this task is *conceptual* rather than *procedural* (Wertheimer 1985; Chronicle, MacGregor, and Ormerod 2004).

3.6.2 The Simulation Model

A simulation model of the case study explained above was developed using SIMUL8 Education Edition 2013 (Simul8 2013) (Figure 2). Subjects in the experimental conditions were given the same simulation model with the exception that participants in the animation condition could watch only the animated display of the model, whereas participants in the statistics condition could access only the statistical outcomes generated from each simulation run.

The animated display of the model can be viewed as a dynamic picture that changes whenever an event occurs. A few examples of possible events in this context are: a patient calls NHS111, a call is transferred to a nurse, a member of staff changes status from idle to busy, and a caller abandons the system. In this way, the logic of the system was communicated in a manner that was visual. In addition, some real time results of the model such as the size of a queue, and the number of the abandoned, overestimated and completed calls were shown on screen. At the end of a run, the performance (i.e. total cost and mean time in system) of each scenario was also reported on screen.

The statistical outcomes generated from each simulation run can be considered as a numerical description of the system's performance. In particular, the following results for each scenario were given in a new window: the total cost; the cost and the rates of abandoned calls and overestimated referrals; the mean time in system and a histogram of time in system; the mean waiting time and a histogram of waiting times for each type of staff.

3.6.3. Instruction Sheets

Students were provided with written instructions about how to approach the problem. The instructions differed subject to the condition the participants were assigned to. In particular, for the experimental conditions, the instructions focused on how to use the simulation model (i.e. set-up of scenarios and running the model). For the control group, the instructions explained what aspects of the students' problem solving approach they needed to document.

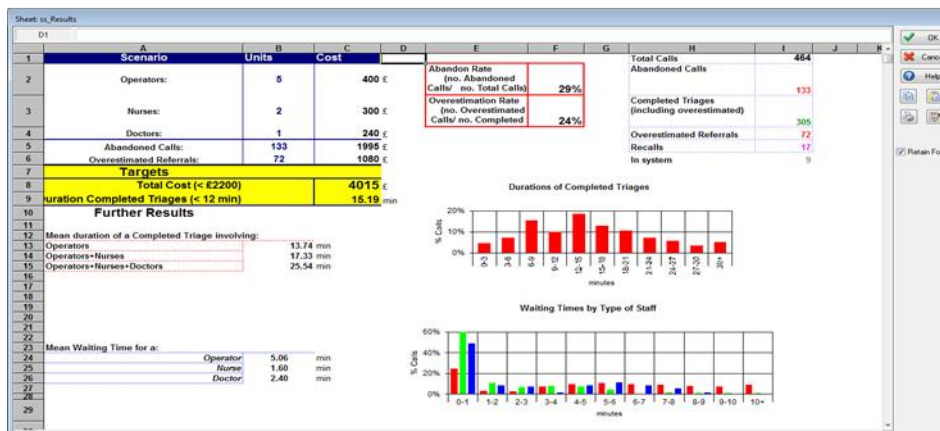
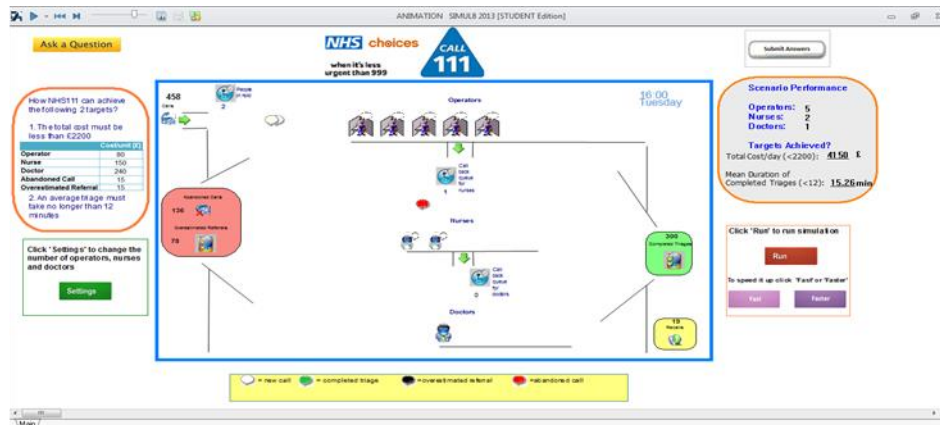


Figure 2. The animated display of the simulation model (above) and the statistical outcomes generated from a simulation run (below).

4 Results

This section discusses the results of the statistical analyses conducted with respect to the two hypotheses of this study. For these analyses we run a number of chi-square tests to draw statistical inferences about the proportion of insights in each group. With respect to the populations that the data are drawn from, we use a test of independence (one population) or a test of homogeneity (two populations). Due to the relatively small sample size and unknown population frequencies, we perform tests by incorporating Yates' correction into the expression for χ^2 . Fleiss, Levin and Paik (2003, 57-58) explain that this adjustment should not only be used to correct the error introduced by using the continuous chi-squared distribution to approximate the discrete distribution of observed binomial frequencies in a two-way contingency table, but also to bring into closer agreement the exact probabilities that are used when the population probabilities are unknown.

4.1 Hypothesis 1: Insights from the Simulation Models (Simulation > No-Simulation)

The solution rates for the three groups are as follows: animation 48% (12/25); statistics 59% (13/22); control 25% (5/20). We note that among the non-solvers there are two statistics participants and six animation participants who managed to generate a new idea (i.e. employing no doctors) that is relevant to finding the solution, but is not the solution to the problem (i.e. false insights). Also, we had to include among the non-solvers, three statistics participants who solved the problem, but provided answers to the post-session questionnaire that did not correspond to the solution. It seems that they may have either changed their mind about the solution to the problem or not realised that one of the scenarios they ran actually solved the problem. An overall 2 (simulation/no simulation) x 2 (solution/no solution) chi-square test of homogeneity with Yate's correction was conducted. The test reveals that the probability of solving the problem among the participants who did not use a simulation model is significantly lower when compared with the solution rate of participants who used the simulation model. Specifically, the one-tailed p -value associated with the $\chi^2(1, N = 67)$ test was below 0.05 ($p = 0.032$). The results suggest there is some association between the use of simulation models and problem solving (Phi = 0.259) (Cohen 1988).

Based on the operationalisation of the concept of insight in the simulation context (see section 2.2), we further investigate whether the above solution rates indicate the occurrence of insight. In detail, initial scenario and pre-session answers were coded by the one of the authors (AG) to determine if a participant experienced an impasse or directly generated the solution after reading the case study. This analysis reveals that two solvers in the control group did not face an impasse as they solved the problem in their first attempt. In addition, post-session answers were coded to determine whether a participant, after reaching the solution, had a clear understanding of the reasons their suggested scenario solved the problem. Five solvers in the animation group and two solvers in the statistics group were found unable to justify their suggested solution. As a result, the solution rates that indicate occurrence of insight for the three groups are counted as those participants who did not immediately solve the problem, but did find the solution, and could justify their answer at the end of the task. The solution rates indicating occurrence of insight for each group are given in Table 3.

Table 3. Solution rates that indicate occurrence of insight for the experimental conditions and control groups.

	Experimental Conditions	Control Group	Significant Difference?
	Simulation Users	Non-Simulation Users	
Solution rates indicating occurrence of insight	38% (18/47)		No ($p = 0.055$)
	<i>'Animation'</i> 24% (7/25)	15% (3/20)	No ($p = 0.25$)
	<i>'Statistics'</i> 50% (11/22)		Yes ($p = 0.02$)

To test the assumption that insights are generated more frequently when a simulation model is used, we adopt the same chi-square test as above comparing the solution rates found in Table 3 for the experimental conditions group and control group. The results reveal that the difference between simulation users and non-simulation users in proportions of solutions indicating occurrence of insight are not quite statistically significant. In particular, the one-tailed p -value related to the $\chi^2(1, N = 67)$ test is just above 0.05 ($p = 0.055$).

As the solution rates that indicate occurrence of insight differ between animation and statistics conditions (see Table 3), two follow up 2×2 chi-square analyses of homogeneity with Yate's correction were conducted to compare insight frequency of the control group against the animation and statistics conditions separately. Note that because two hypotheses had to be tested on one set of data, the chance of obtaining a false-positive result (type-I error) is 9.75 per cent. Accordingly, we use the Bonferroni correction to adjust the critical value (α) for each hypothesis to 0.025 to counterbalance this risk (Bland and Altman 1995). The results reveal that participants in the statistics condition have a significantly higher solution rate indicating greater occurrence of insight than participants in the control group. More specifically, the p -value associated with the $\chi^2(1, N = 42)$ test is below the adjusted α significance level 0.025 ($p = 0.02$). The results suggest a medium association between the use of statistical outcomes of the simulation model and insight generation ($\Phi = 0.370$). In addition, no difference in solution rates indicating occurrence of insight is found between participants in the animation condition and control group. The one-tailed p -value related to $\chi^2(1, N = 45)$ test is above the adjusted α significance level 0.025 ($p = 0.25$).

To summarise, the task results provide some support for hypothesis 1. More specifically, students who used the simulation model solved the task more often than students who did not. The proportion of solvers who indicate occurrence of insight is significantly greater in the statistics group than in the control group. However, the relevant proportion of solvers who indicate occurrence of insight between the animation group and the control group do not differ significantly.

4.2 Hypothesis 2: Insights from the Model’s Animation and Statistical Outcomes

(Animation ≠ Statistics)

In the following subsections we look at differences in task performance and problem-solving patterns between the participants in the animation and statistics conditions who solved the task by indicating occurrence of insight.

4.2.1 Task Results

Participants in the statistics group have higher solution rates indicating occurrence of insight (50%) than in the animation group (24%) (Table 3). To test whether there is a significant association between the model representation and occurrence of insight, a 2 (animation/statistics) x 2 (insight/no insight) chi-square test of independence with Yate’s correction was conducted comparing the solution rates found in Table 3 for the animation and statistics conditions. The results reveal that there is not enough evidence to confirm an association between the model representation and occurrence of insight; the two-tailed p -value related to the χ^2 (1, $n = 47$) test is above 0.05 ($p = 0.21$). So there is no significant difference in the solution rates that indicate occurrence of insight between the animation and statistics groups.

We also test whether demographic characteristics influenced the occurrence of insight. The demographic characteristics of the participants in the experimental conditions who indicated occurrence of insight are presented in Table 4. There were no statistically significant differences between all the simulation users ($n=47$, Table 2) and those who indicated experience of insight ($n=18$) in terms of gender, prior experience with the context of the case study (NHS111 or a similar service), degree year and the number of modules taken with a quantitative content.

Table 4. Characteristics of the participants who indicated experience of insight for the experimental conditions.

	'Animation' Condition		'Statistics' Condition		Significant Difference Simulation users ($n=47$, Table 2) to simulation users with insight ($n=18$)?
	$n = 7$		$n = 11$		
	n	%	n	%	
Gender:					No ($p = 1.000$) ^a
Male	4	57%	6	55%	
Female	3	43%	5	45%	
Prior Experience:					No ($p = 0.663$) ^a
Yes	0	0%	1	9%	
No	7	100%	10	91%	
Degree Year:					No ($p = 0.241$) ^a
1	2	29%	8	73%	
2	5	71%	3	27%	
Quantitative Modules:					No ($p = 0.267$) ^a
≤ 3	2	29%	6	55%	
4-6	5	71%	5	45%	

^aDifferences between the groups were tested with Fisher exact tests.

4.2.2 Problem Solving Patterns Results

To obtain a visual comparison of the problem-solving approach of each participant, the possible scenarios for the problem used in the experiment are categorised into 5 categories (Table 5). We display in a scatter plot the categories of scenarios each participant ran against the time each scenario was set up. The categorisation of scenarios was developed in parallel with the creation of the case study and the simulation model. Category 1 and Category 2 consist of scenarios in which all types of staff are used. The scenarios of Category 1 move participants away from the solution, whereas scenarios of category 2 bring them closer to the solution. The scenarios of Category 3, 4 and 5 indicate occurrence of insight. Although the scenarios of Category 3 (a system with no clinical experts) involve overcoming self-imposed constraints, they are false insights as they show a misunderstanding of the problem (Isaak and Just 1995); the system cannot run without experts. Category 5 is a subgroup of Category 4 and includes the scenarios that are deemed to solve the problem. Examples of participants' problem solving patterns are shown in Figure 3.

Table 5. The possible scenarios for the NHS111 problem separated into 5 categories.

Category	Scenario Description
1	Operators \geq Doctors + Nurses
2	Operators $<$ Doctors + Nurses
3	A service with no Experts (i.e. No Doctors or/and no Nurses)
4	A service with only Experts (i.e. No Operators)
5	A system with only 5 Doctors OR a system with 5 Nurses and up to 2 Doctors

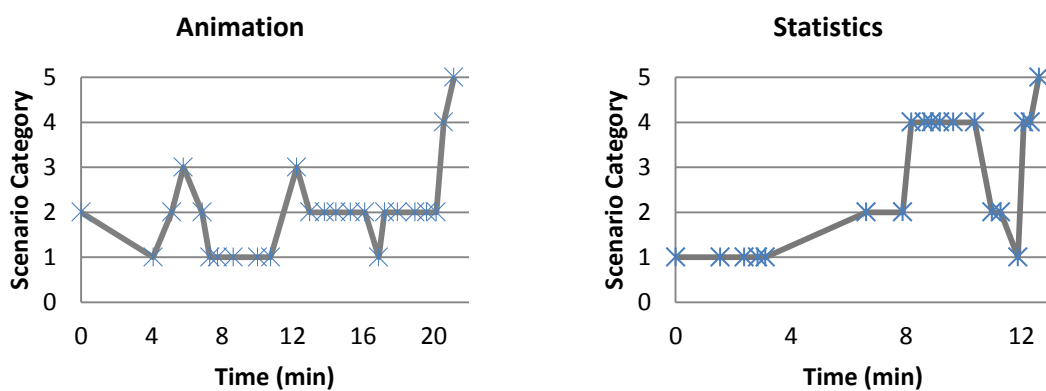


Figure 3. Two examples of problem-solving patterns of a participant who solved the problem in the animation condition (left) and in the statistics condition (right).

In order to increase our understanding of the problem solving patterns of the participants who indicated occurrence of insight in the animation ($n = 7$) and statistics ($n = 11$) conditions, we split the problem solving patterns into two parts: *before- and after-breaking the impasse*. By *before-breaking the impasse* we refer to the part of a problem solving pattern up to a participant overcoming self-imposed constraints; that is, running scenarios in Category 1 or 2 (Table 5). *After-breaking the impasse* is the segment of the problem solving pattern from a participant breaking the impasse (running a scenario in Category 3 or 4) up to first solving the task (running for the first time a scenario in Category 5). If breaking the impasse and solving the task occur together, then *after-breaking the impasse* consists of just one step; a scenario in Category 5. For each part, we run descriptive statistics on two key features: *duration* and *number of scenarios* (Table 6). We also measure the total *task duration* and the *scenario rate*; that is, the average number of scenarios a participant ran per minute of experimentation. The median as well as the lower and upper quartiles for the above measures for each group can be found in Table 6.

Table 6. Analysis of the problem solving patterns with insight, split into before- and after-breaking the impasse, and task information.

	'Animation' $n = 7$		'Statistics' $n = 11$	
	Median	(Lower - Upper Quartiles)	Median	(Lower - Upper Quartiles)
<u>Before - breaking the impasse</u>				
<i>Duration (min)</i>	9.42	(5.80 - 12.07)	9.05	(3.37 - 10.67)
<i>Number of Scenarios</i>	12	(5 – 25)	11	(6 - 32)
<u>After - breaking the impasse</u>				
<i>Duration (min)</i>	11.28	(0.93 - 12.88)	2.33	(1.52 - 4.78)
<i>Number of Scenarios</i>	21	(3 – 27)	8	(3 – 15)
<u>Task</u>				
<i>Duration (min)</i>	20.70	(12.06 - 22.45)	11.38	(8.48 - 15.20)
<i>Scenario rate</i>	1.66	(1.52 – 3.19)	2.36	(1.81 – 3.69)

The results show that breaking the impasse (*before-breaking the impasse*, Table 6) typically occurs at similar times and number of scenarios run for both the animation and statistics participants, and with a similar variation in the results. After the impasse is broken, few animation participants require less time than statistics participants to reach the solution (lower quartiles *after-breaking the impasse duration*), and the majority of the animation participants need about five times as long as the statistics participants require to solve the problem (medians *after-breaking the impasse duration*). Accordingly, they need to run more scenarios than the statistics participants with higher variation in the results (upper and lower quartiles *after-breaking the impasse, number of scenarios*).

Consequently, the *task duration* of the animation participants tends to be 9.32 more minutes than for statistics participants, and with higher variation in the outcome (median *task duration* animation = 20.70, median *task duration* statistics = 11.38). We ran a Mann-Witney's U-test to evaluate the difference in *task duration* between the animation and statistics participants. The results suggest that the *task duration* of the statistics participants is significantly lower than that of the animation participants ($p = 0.035 < 0.05$, $r = 0.50$). Also, taking account of the *scenario rates*, the above results suggest that the statistics participants solve the task more rapidly than the animation participants.

In an attempt to identify further differences in how the animation and statistics participants generate insights, we determine the most common problem solving pattern for each experimental condition (Figure 4). More specifically, for each minute of experimentation, we identify the category of scenario that is run most often by the animation and statistics participants. These modal values are plotted sequentially on Figure 4. This shows that the majority of the animation and statistics participants initially thought that assigning more operators was an effective strategy (scenario Category 1). They then moved to reducing the number of operators (scenario Category 2). Both groups continued for 9 minutes before the implicit imposed constraint of 'assigning all types of staff' was broken first by the majority of the statistics group.

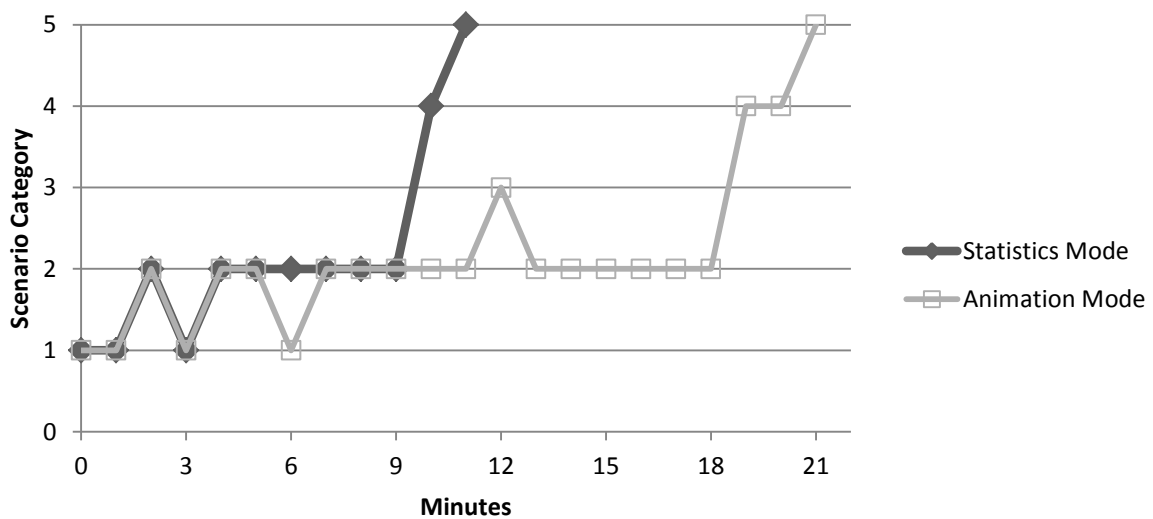


Figure 4. The modal problem solving patterns with insight for participants in the animation and statistics conditions.

Whilst most statistics participants solved the task (scenario Category 5) soon after breaking the impasse (scenario Category 4), this was not the case for the animation participants. The majority of the animation participants first attempted scenarios without experts (scenario Category 3). Although this involves breaking the impasse, it is a false insight because it indicates a misinterpretation of the cause of the problem (Isaak and Just 1995); the system cannot work without experts. Accordingly, straight after breaking the impasse, the majority of the animation participants returned to the

previous strategy (scenario Category 2), instead of solving the task. Eventually, the majority of the animation participants found the solution to the problem after 20 minutes of experimentation. The above results show a very different problem solving pattern for the majority of the statistics versus animation participants, with the statistics participants generating insights much more rapidly.

In order to further understand how participants in the experimental conditions generated insights, we compare the problem solving patterns of the simulation users who indicated occurrence of insight (n = 18, Table 3) to the participants in the experimental conditions who did not achieve insight (n = 29), irrespective of the condition they were allocated to. We run descriptive statistics (median, lower and upper quartiles) on two key features: *task duration* and *number of scenarios*. This analysis is presented in Table 7.

Table 7. Analysis of the problem solving patterns of the participants in the experimental conditions who indicated occurrence of insight and who did not.

	Insight n = 18		'Non-Insight' n = 29	
	Median	(Lower – Upper Quartiles)	Median	(Lower – Upper Quartiles)
Task				
<i>Duration (min)</i>	12.10	(8.33 - 19.13)	25.56	(21.20 - 28.60)
<i>Number of Scenarios</i>	29	(12 – 44)	46	(35 – 63)

The results in Table 7 show that the participants who did not indicate occurrence of insight tend to spend about twice as much time on the task as the participants who generated insight (medians, *duration*), with lower variation in the results (upper and lower quartiles, *duration*). Not surprisingly, they typically ran 17 more scenarios than the subjects who generated insight with higher variation in the results (median *number of scenarios* insight = 29, median *number of scenarios* non-insight = 46). The results of Mann-Witney’s U-tests suggest that both *task duration* and *number of scenarios* for the participants who indicated occurrence of insight are significantly lower than that of the participants who did not generate insight (*task duration* $p < 0.001$, $r = 0.63$; *number of scenarios* $p = 0.009 < 0.05$, $r = 0.38$). These results indicate that the participants who do not indicate occurrence of insight do not manage to change their understanding of the cause of the NHS111 problem and therefore continue playing around with the model in a state of impasse, instead of solving the task.

In summary, the results of the above analyses provide evidence that using the statistical results generated from simulation runs is more helpful in insight generation than watching the animation of the model. Although there is no significant difference in solution rates that indicate occurrence of insight between the groups, most of the statistics participants generated insights more rapidly than the animation participants. As such, we find support for the second hypothesis that the contribution of animation to the process of insight generation differs from the contribution of statistical outcomes. The results, however, are quite the opposite of the expectation set by the work of Chau and Bell (1994) and Bell and O'Keefe (1995) in which it is claimed that the animated display is the most helpful. Based on the results of the above analyses and in relation to the possible explanations about the mental mechanisms of insight generation introduced in section 2.1, it seems that in our case study insights emerge after a participant overcomes implicitly imposed constraints and changes his/her understanding of the cause of the problem.

5 Discussion

Having presented the results of the experiments, this section discusses the conclusions with respect to the hypotheses of the research, considers the limitations of the study and provides suggestions for future work.

5.1 Conclusions about Hypotheses

The first hypothesis seeks to determine whether insights are generated more frequently when a simulation (DES) model is used. Our experimental results provide some evidence to support this hypothesis. The simulation users (experimental conditions group) did solve the NHS111 problem significantly more frequently than the non-simulation users (control group). However, when comparing the solution rates that indicate the occurrence of insight between the two groups, the statistical result (p -value = 0.055) is not significant at the accepted level (α = 0.05), but it does provide weak evidence for an effect of the simulation model on insight generation. When the simulation users are split into their two sub-groups (animation and statistics), we find that the statistics group have a significantly greater solution rate with insight than the non-simulation users; this is not the case for the animation group. This evidence corresponds with recent research which has found that experimentation with a simulation model supports learning and understanding (Monks, Robinson, and Kotiadis 2014).

The second research hypothesis asks whether the contribution of animation to the process of insight differs from the contribution of the statistical results generated from a simulation run. In examining the solution rates that indicate occurrence of insight, the experimental results do not provide

sufficient evidence that either group (animation or statistics) generated insights more frequently than the other. However, when the problem-solving patterns for each group are analysed, the results suggest that using statistical outcomes generates insights more rapidly than the use of animation. This finding is in contrast with previous research that highlights the value of simulation animation over statistical outcomes (Chau and Bell 1994, Bell and O'Keefe 1995).

Although the participants in the animation condition broke the impasse at similar times to the participants in the statistics group, they appear to first have a false insight (Isaak and Just 1995). That is, the animation participants tended to draw inferences about the NHS111 problem in an illogical fashion. In particular, when they realised that the self-imposed constraint of 'all types of staff must be used' can be broken, they first considered eliminating expert rather than inexperienced staff from the system. As a result, breaking the impasse was not necessarily associated with generating insights for the animation participants. We note that this result may be specific to this task. A different outcome might have been produced if the characteristics of the case study and task were different.

The above findings are in alignment with recent experimental studies in the system dynamics field which have shown that people employ a number of judgmental biases and heuristics in complex situations (Sterman 1989; Sterman 1994; Langley and Morecroft 2004; Monxes 2004; Sterman and Booth Sweeney 2007; Cronin, Gonzalez, and Sterman 2009). Researchers explain that people tend to systematically misperceive the behaviour of even the simplest dynamic systems and hence make use of counterproductive heuristics. Taking into account the level of difficulty of the case study used in this experiment (i.e. insight problem), it is not surprising that some subjects misunderstood the true cause of the NHS111 problem, and therefore first considered eliminating expert rather than inexperienced staff from the system. Further, Hämäläinen, Luoma and Saarinen (2013) argue that heuristics may also be triggered by a number of factors that camouflage the true system's operation. Based on this view, it may be that in our experiment the animated display did not reveal the behaviour of the system as clearly as the statistical results.

It is also possible that participants who used the animation of the simulation model mainly set up scenarios by activating System 1 thinking (Kahneman 2011). Watching the animated display of a simulation model lends itself to an intuitive problem solving approach without the need, or even possibility, of analysis. As a result, participants in the animation condition probably ran scenarios based on their intuition, by activating System 1. However, making decisions based on intuition encompasses the risk of overestimating benefits and underestimating costs. For true insights to emerge it is necessary to consciously understand the cause of a problem (Dane and Pratt 2007, Akinici and Sadler-Smith 2012).

The descriptive statistics from the experiments (Table 6) show that for both animation and statistics participants, the insight generation and solving the problem did not occur together; problem solving occurred later. This finding corresponds with Schooler, Fallshore and Fiore's (1995) definition and previous research that show that having an Aha! moment may happen suddenly, but solving the problem might require some additional steps (MacGregor, Ormerod, and Chronicle 2001). The problem solving patterns of both the animation and statistics participants show that subjects tend to use an anchoring approach where they stick to one type of scenarios which was only disregarded when a better solution was found (Kanheman 2011). This result confirms previous experimental research which shows that simulation users tend to set up scenarios which confirm rather than challenge their thinking when experimenting with a model; this is referred to as 'confirmation bias' (Bell and O'Keefe 1995; Monks, Robinson, and Kotiadis 2014).

5.2 Limitations of the Study

Despite the careful design of the experiment, this study is subject to some limitations. These relate to the participants in the study, the task and case study, and the assessment of participants' understanding.

In terms of the study participants, they were not randomly assigned between the experimental conditions and control groups. It is known that the control group are more experienced than the experimental conditions group. As such, we would expect the control group to perform better on the task than the experimental conditions group. However, the less experienced participants (simulation users) generate insights at least as frequently as the more experienced participants (non-simulation users); and more frequently in the case of the statistics group. So, given the direction of the results and the difference in the experience of the groups, this lends greater weight to the benefits of simulation in generating insight.

A second issue with the participants relates to the use of students and whether their performance would differ from that of managers. Of course, the choice of students was driven by convenience. It would be extremely difficult to obtain the participation of sufficient managers and there would be far greater concerns about the range of experience among managers. This concern is mitigated by the findings from the experimental study of Bakken, Gould and Kim (1992). They found that students performed better than managers on their simulation based experimental task because the students did not possess prior knowledge and experience which prevented them from freely experimenting with the model. So, students, whose minds are *tabula rasa* (blank slates) compared to managers, are expected to be better than managers at generating insights from simulation models.

A final concern regarding the participants is the small sample size for solution rates indicating insight (animation $n = 7$, statistics $n = 11$). As a result, it is difficult to determine whether the distributions for task duration of the animation and statistics conditions have the same shape. Some care should be taken, therefore, with interpreting the inferential statistics concerning task duration.

The characteristics of the task and the case study for this experiment might have had an impact on the findings. It may be that the characteristics of the NHS111 problem lend themselves better to being understood through statistical outcomes than a simulation animation. Indeed, previous research shows that the appropriateness of the display depends on the characteristics of the task (Dickson 1988, Jarvenpaa 1989, O'Keefe and Pitt 1991). A different task might be better translated into pictures than in numbers. As such, the results may be specific to this task and cannot be generalised.

In a similar vein, a different animated view of the NHS111 model and a different display of the statistical outcomes generated from the model might have elicited different levels of performance by the participants. It is important to take account of display preference needs (Polys, Bowman, and North 2011) and human-computer interface design principles (Howie et al. 2000). That said, the 13 participants in pilot tests commented that the animated display and output statistics are appropriate for the given task.

Consideration must be given to the accuracy of the measures of participants' understanding. Specifically, we evaluate participant understanding through pre- and post- session questionnaires which asked for written explanations. These answers may not have fully reflected the participants' understanding of the problem. For instance, we may not have identified a participant who obtained insight simply because the written answers to the post-session questionnaire were incomplete. If the participants were asked to answer the same questions verbally, their answers may have been more descriptive. There is also subjectivity involved in marking the participants' answers to determine whether they were indicating an insight. To mitigate this effect, the first author made three independent evaluations of participants' answers, approximately three weeks apart, following the same coding system. The consistency in scoring was 91%, with differences resolved by discussion with the two other authors.

5.3 Further Work

This study demonstrates that there are some differences between the use of animation and statistics in generating insight from using DES models. In particular, the reporting of statistical

outcomes appears to be more effective than the use of animation. This is a surprising result given past claims and research studies. Certainly, these are relationships that should be explored further.

Future work could consider using more than one case studies (insight problems), where participants are randomly assigned to different tasks or experimental design. It would also be useful to explore how models can be effectively designed to generate insight by giving participants alternative visual representations (animation and statistics) of the same model. It would also be useful to perform a similar experiment with managers, although it would be much more difficult to obtain a reasonable sample size.

To mitigate the subjectivity in the measurement of insight, an additional measure for future work could be transfer of learning. Participants would be asked to transfer the knowledge obtained from one insight problem to another problem in a different context, but with a similar structure (Bakken, Gould and Kim 1992). Future work could also consider studying the mental process that simulation users follow to generate insights. It would be interesting to know how people identify and overcome behavioural issues and hence proceed from the impasse phase to illumination.

Furthermore, experimental methods could be used as well as field studies to study the effectiveness of other aspects of a simulation intervention in insight generation, such as the role of the modeller or facilitator, problem formulation and the involvement of stakeholders from different departments. This would entail observing the development and use of simulation models, and recording the occurrence of both impasses and insights.

6. Conclusions

This research explores the role of DES models in generating insights. A controlled quasi-experiment is employed, using a between-groups design. One independent variable is manipulated: the features of the simulation model. Participants work on a task either using only a simulation animation or only the statistical results of the model. Meanwhile, a control group work on the task without the use of a model.

To our knowledge, ours is the only experimental study attempted so far that has tried to investigate the use of simulation models in insight generation. To an extent, the results of the experiment support the claim that insights are generated more frequently when a simulation model is used. While the differences in insight frequencies between the statistics condition and the control group are statistically significant, those between the animation and control group are not. Although there is no evidence that insights are generated more frequently when participants used the statistical

outcomes rather than the animation of the simulation model, we find that the use of statistical outcomes reduces the emergence of false insights, and that has a direct impact on task duration.

Despite some limitations in this experimental study, the results of the research provide empirical evidence about the value of DES in insight generation. We also provide workers in behavioural operational research with an approach for studying insight using laboratory based experiments. Future work should continue to explore insight generation from models, and from the wider activity of developing and using models.

Acknowledgments

The authors acknowledge the financial support and advisory input of SIMUL8 Corporation as well as the feedback of Dr. Khayat on the case study. The literature review of this paper is based on Gogi, A., A. A. Tako, and S. Robinson. 2014. "Generating Insights: The Effectiveness of Simulation Models in Creative Problem Solving". In *Proceedings of the Operational Research Society Simulation Workshop 2014*, edited by B.Tjahjono, C. Heavey, S. Onggo, and D-J. van der Zee, 133-142. The Operational Research Society, Birmingham, UK; Gogi, A., A. A. Tako, and S. Robinson. 2014. "A Preliminary Study on the Role of Simulation Models in Generating Insights". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. D. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller. IEEE, Savannah, Georgia.

References

- Abeler, J., and D. Nosenzo. 2013. "Self-Selection into Economics Experiments Is Driven by Monetary Rewards." Discussion Papers 7374, Institute for the Study of Labor (IZA), Bonn, Germany.
- Agresti, A. 1990. *Categorical Data Analysis*. New York: John Wiley & Sons.
- Akinci, C., and E. Sadler-Smith. 2012. "Intuition in Management Research: A Historical Review." *International Journal of Management Reviews* 14 (1): 104-122.
- Argyris, C., and D. A. Schön. 1996. *Organisational Learning II: Theory Method and Practice*. Addison Wesley.
- Bakken, B., J. Gould, and D. Kim. 1992. "Experimentation in Learning Organizations: a Management Flight Simulator Approach." *European Journal of Operational Research* 59 (1): 167-182.
- Bayer, S., T. Bolt, S. Brailsford, and M. Kapsali. 2014. "Models as Interfaces". In *Discrete-Event Simulation and System Dynamics for Management Decision Making*, edited by S. Brailsford, L. Churilov, and B. Dangerfield, 125-139. Chichester: John Wiley & Sons.
- Bell, P. C., and R. M. O'Keefe. 1987. "Visual Interactive Simulation — History, Recent Developments, and Major issues." *Simulation* 49 (3): 109-116.
- Bell, P. C., and R. M. O'Keefe. 1995. "An Experimental Investigation into the Efficacy of Visual Interactive Simulation." *Management Science* 41: 1018-1038.
- Belton, V. and M. Elder. 1994. "Decision Support Systems: Learning from Visual Interactive Modelling." *Decision Support Systems* 12: 355-364.
- Bland, J. M., and D. G. Altman. 1995. "Multiple Significance Tests: The Bonferroni Method." *British Medical Journal* 310: 170.

- Bonner, S. E., and G. B. Sprinkle. 2002. "The Effects of Monetary Incentives on Effort and Task Performance: Theories, Evidence, and a Framework for Research." *Accounting, Organizations and Society* 27 (4): 303-345.
- Campbell, D., and J. Stanley. 1963. *Experimental and Quasi-experimental Designs for Research*. Chicago, IL: Rand-McNally.
- Carpenter, M. L., Jr. K. W. Bauer, T. F. Schuppe, and M. A. Vidulich. 1993. "Animation: What's Essential for Effective Communication of Military Simulation Model Operation?" In *Proceeding of the 1993 Winter Simulation Conference* edited by G. W. Evans, M. Mollaghasemi, E. C. Rusell and W. E. Biles, 1081-1088. IEEE, Los Angeles, California, doi: 10.1109/WSC.1993.718364.
- Chau, P. and P. C. Bell. 1994. "Decision Support for a New Production Plant Design Using Visual Interactive Simulation." *Journal of Operational Research Society* 45: 1273-1284.
- Chronicle, E. P., J. N. MacGregor, and T. C. Ormerod. 2004. "What Makes an Insight Problem? The Roles of Heuristics, Goal Conception, and Solution Recoding in Knowledge-Lean Problems" *Journal of Experimental Psychology: Learning, Memory and Cognition* 30 (1): 14-27.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd Edition, Hillsdale: L. Erlbaum Associates.
- Cronin, M. A., C. Gonzalez, and J. D. Serman. 2009. "Why Don't Well-Educated Adults Understand Accumulation? A Challenge to Researchers, Educators, and Citizens." *Organizational Behavior and Human Decision Processes* 108 (1): 116-130.
- Crookes, J.G. 1982 "Simulation in 1981." *European Journal of Operation Research* 9: 1-7.
- Dane, E., and M. Pratt. 2007. "Exploring Intuition and its Role in Managerial Decision Making." *Academy of Management Review* 32 (1): 33-54.
- Davidson, J. E. 1995. "The Suddenness of Insight." In *The Nature of Insight*, edited by R. J. Sternberg and J. E. Davidson, 125-155. Cambridge, MA: The MIT Press.
- de Vreede, G-J., and A. Verbraeck. 1996. "Animating Organizational Processes Insight Eases Change" *Simulation Practice and Theory* 4: 245-263.
- den Hengst, M., G-J. de Vreede, and R. Maghnoouji. 2007. "Using Soft OR Principles for Collaborative Simulation: A Case Study in the Dutch Airline Industry." *Journal of the Operational Research Society* 58: 669-682.
- Dickson, G. W. 1988. "Graphics and Managerial Decision Making: Research Based Guidelines." *Communications of the ACM* 31 (6): 764-774.
- Elder, M. D. 1992. *Visual Interactive Modelling: Some Guidelines for its Implementation and Some Aspects of its Potential Impact on Operational Research*. PhD Thesis, Department of Management Science, University of Strathclyde, Edinburgh.
- Fleiss, J. L., B. Levin, and M. C. Paik. 2003. *Statistical Methods for Rates and Proportions*. 3rd Edition, New Jersey: John Wiley and Sons.
- Fletcher, A., D. Halsall, D. Huxham and D. Worthington. 2007. "The DH Accident and Emergency Department Model: A National Generic Model Used Locally." *Journal of the Operational Research Society* 58 (12): 1554-1562.
- Fone, D., et al. 2003. "Systematic Review of the Use and Value of Computer Simulation Modelling in Population Health and Health Care Delivery." *Journal of Public Health Medicine* 25 (4): 325-335, doi: 10.1093/pubmed/fdg075.
- Gogi, A., A. A. Tako, and S. Robinson. 2014. "Generating Insights: The Effectiveness of Simulation Models in Creative Problem Solving". In *Proceedings of the Operational Research Society Simulation Workshop 2014*, edited by B.Tjahjono, C. Heavey, S. Onggo, and D-J. van der Zee, 133-142. The Operational Research Society, Birmingham, UK
- Griffiths, D., J. E. Williams, and R. M. Wood. 2013. "Modelling Activities at a Neurological Rehabilitation Unit." *European Journal of Operational Research* 226 (2): 301-312.
- Günal, M. M. and M. Pidd. 2010. "Discrete Event Simulation for Performance Modelling in Health Care: A Review of the Literature." *Journal of Simulation* 4: 42-51.

- Hämäläinen, R. P., J. Luoma, and E. Saarinen. 2013. "On the Importance of Behavioural Operational Research: The Case of Understanding and Communicating about Dynamic Systems" *European Journal of Operational Research* 228 (3): 623-634.
- Hsiao, N., and G. P. Richardson. 1999. "In Search of Theories of Dynamic Decision Making: A Literature Review". In *the 17th International Conference of the System Dynamics Society*. New Zealand: Wellington. <http://www.systemdynamics.org/conferences/1999/PAPERS/PLEN16.PDF>
- Howie, E., S. Sy, L. Ford, and K. J. Vicente. 2000. "Human-computer Interface Design Can Reduce Misperceptions of Feedback." *System Dynamics Review* 16(3): 151-171.
- Hurriion, R. D. 1986. "Invited Review: Visual Interactive Modelling" *European Journal of the Operational Research* 54: 281-287, doi: 10.1016/0377-2217(86)90293-6.
- Isaak, M. I., and M. A. Just. 1995. "Constraints on Thinking in Insight and Invention." In *The Nature of Insight*, edited by R. J. Sternberg and J.E. Davidson, 281-325. Cambridge, MA: MIT Press.
- Jarvenpaa, S. 1989. "The Effect of Task and Graphical Format on Information Processing Strategies." *Management Science* 35 (3): 285- 303.
- Jun, J. B., S. H. Jacobson, and J. R. Swisher. 1999. "Applications of Discrete Event Simulation in Health Care Clinics: a Survey." *Journal of the Operational Research Society* 50: 109–123.
- Kahneman, D. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kounios, J., and M. Beeman. 2009. "The Aha! Moment: the Cognitive Neuroscience of Insight." *Current Directions in Psychological Science* 18 (4): 210-216.
- Langley, P. A., and J. D. W. Morecroft. 2004. "Performance and Learning in a Simulation of Oil Industry Dynamics." *European Journal of Operational Research* 155 (3): 715-732.
- Lehaney, B., S. A. Clarke, and R. J. Paul. 1999. "A Case of an Intervention in an Outpatients Department." *Journal of the Operational Research Society* 50: 877-891.
- Luehrmann, M. A. H., and D. L. Byrkett. 1989. "A Pilot Study of the Impact of Animation and Decision-Making." *Simulation* 53: 153-158.
- Macal, C. M. 2001. "Introduction to Special Issue on Simulation and Visualization." *Simulation* 77: 90–92.
- MacGregor, J. N., T. C. Ormerod, and E. P. Chronicle. 2001. "Information Processing and Insight: A Process Model of Performance on the 9-Dot Problem and Related Problems" *Journal of Experimental Psychology: Learning, memory and Cognition* 27 (1): 176-201.
- Maier, N. R. F. 1940. "The Behavior Mechanisms Concerned with Problem Solving." *Psychological Review* 47 (1): 43-58.
- Mayer, R. E. 2010. "Problem Solving and Reasoning." In *International Encyclopedia of Education*, edited by P. Peterson, E. Baker, and B. McGaw, 273-278. 3rd Edition, Oxford: Elsevier.
- Metcalfe, J., and D. Wiebe. 1987. "Intuition in Insight and Noninsight Problem Solving." *Memory & Cognition* 15 (3): 238-246.
- Monks, T., S. Robinson, and K. Kotiadis. 2014. "Learning from Discrete-Event Simulation: Exploring the High Involvement Hypothesis." *European Journal of Operational Research*, 235 (1): 195-205.
- Monxes, E. 2004. "Misperceptions of Basic Dynamics: The Case of Renewable Resource Management." *System Dynamics Review* 20: 139-162.
- O'Kane, J. F. 2004. "Simulating Production Performance: Cross Case Analysis and Policy Implications", *Industrial Management & Data Systems* 104 (4): 309 – 321.
- O'Keefe, R. M. and I. L. Pitt. 1991. "Interaction with a Visual Interactive Simulation, and the Effect of Cognitive Style." *European Journal of Operational Research* 54: 339-348.
- Parker, D. C. 1991. "Visual Interactive Financial Models: An Overview of Micro-Computer Software Offerings and Discussion of Potential Decision Support." *European Journal of Operational Research* 54: 330-338.
- Paul, R. J., G. M. Giaglis, and V. Hlupic. 1999. "Simulation of Business Processes: A Review." *American Behavioral Scientist* 42: 1551–1576.
- Pidd, M. 2010. "Why Modelling and Model Use Matter." *Journal of the Operational Research Society* 61: 14-24.

- Polys, N. F., D. A. Bowman, and C. North. 2011. "The Role of Depth and Gestalt Cues in Information-Rich Virtual Environments." *International Journal of Human Computer Studies* 69: 30-51.
- Powers, M.J., S. M. Sanchez, and T.W. Lucas. 2012. "The Exponential Expansion of Simulation in Research" In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 1552-1563. IEEE, Berlin, Germany.
- Proudlove, N. C., S. Black, and A. Fletcher. 2007. "OR and the Challenge to Improve the NHS: Modelling for Insight and Improvement in In-Patient Flows." *Journal of the Operational Research Society* 58 (2): 145–158.
- Richardson, G. P., D.F. Andersen, T. A. Maxwell, and T. R. Stewart. 1994. "Foundations of Mental Model Research." In *Proceedings of the International System Dynamics Conference 1994*, Stirling, UK.
- Richels, R. 1981. "Building Good Models Is Not Enough." *Interfaces* 11 (4): 48-54.
- Robinson, S. 2001. "Soft with a Hard Centre: Discrete-event Simulation in Facilitation." *Journal of the Operational Research Society* 52: 905-915.
- Robinson, S. 2005. "Discrete-Event Simulation: from the Pioneers to the Present, What Next?" *Journal of the Operational Research Society* 56: 619-629.
- Robinson, S. 2008. "Conceptual Modelling for Simulation Part I: Definition and Requirements." *Journal of the Operational Research Society* 59: 278-290.
- Rouwette, E. A. J. A., A. Größler, and J. A. M. Vennix. 2004. "Exploring Influencing Factors on Rationality: A Literature Review of Dynamic Decision-making Studies in System Dynamics." *Systems Research and Behavioral Science* 21(4): 351–370.
- Rouwette, E. A. J. A., J. A. M. Vennix, and A. J. A. Felling. 2009. "On Evaluating the Performance of Problem Structuring Methods: An Attempt at Formulating a Conceptual Model." *Group Decision and Negotiation* 18(6): 567-587.
- Rouwette, E. A. J. A., H. Korzilius, J. A. M. Vennix, and E. Jacobs. 2011. "Modeling as Persuasion: The Impact of Group Model Building on Attitudes and Behaviour." *System Dynamics Review* 27: 1–21.
- Schooler, J. W., M., Fallshore, and S. M. Fiore. 1995. "Epilogue: Putting Insight into Perspective." In *The Nature of Insight*, edited by R. J. Sternberg and J.E. Davidson, 559-587. Cambridge, MA: The MIT Press.
- SIMUL8. 2013. SIMUL8 Simulation Software, www.simul8.com accessed October 2014.
- Smith, S. M. 1995. "Getting Into and Out of Mental Ruts: a Theory of Fixation, Incubation, and Insight." In *The Nature of Insight*, edited by R. J. Sternberg and J.E. Davidson, 229-251. Cambridge, MA: The MIT Press.
- Sterman, J. D. 1989. "Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment." *Management Science* 35 (3): 321-339.
- Sterman, J. D. 1994. "Learning In and About Complex Systems." *System Dynamics Review* 10 (2-3): 291-330.
- Sterman, J. D., and L. Booth Sweeney. 2007. "Understanding Public Complacency about Climate Change: Adults' Mental Models of Climate Change Violate Conservation of Matter." *Climatic Change* 80 (3–4): 213–238.
- Sternberg, R. J. 2009. *Cognitive Psychology*. 5th Edition, Wadsworth: Cengage Learning.
- Taylor, S. J. E., T. Eldabi, G. Riley, R. J. Paul, and M. Pidd. 2009. "Simulation Modelling is 50! Do we Need a Reality Check?" *Journal of the Operational Research Society* 60: S69-S82.
- Thyer, B. A. 2012. *Quasi-Experimental Research Design*. New York: Oxford University Press, Inc.
- van der Zee, D., and J. Slomp. 2009. "Simulation as a Tool for Gaming and Training in Operations Management - A Case Study." *Journal of Simulation* 3 (1): 17-28.
- Weisberg, R. W. 1995. "Prolegomena to Theories of Insight in Problem Solving: A Taxonomy of Problems." In *The Nature of Insight*, edited by R. J. Sternberg and J.E. Davidson, 157-196. Cambridge, MA: The MIT Press.

- Wenzel, S., and U. Jessen. 2001. "The Integration of 3-D Visualization into the Simulation-based Planning Process of Logistics Systems." *Simulation* 77: 114–127.
- Wertheimer, M. 1985. "A Gestalt Perspective on Computer Simulations of Cognitive Processes." *Computers in Human Behavior* 1: 19-33.