Can involving clients in simulation studies help them solve their future problems? A transfer of learning experiment

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

[a] Faculty of Health Sciences, University of Southampton, Southampton SO17 1BJ, UK
[b] School of Business and Economics, Loughborough University, Loughborough, Leicestershire LE11 3TU, UK
[c] School of Computing, University of Kent, Canterbury, Kent CT2 7NZ, UK

ARTICLE INFO

ABSTRACT

It is often stated that involving the client in operational research studies increases conceptual learning about a system which can then be applied repeatedly to other, similar, systems. Our study provides a novel measurement approach for behavioural OR studies that aim to analyse the impact of modelling in long term problem solving and decision making. In particular, our approach is the first to operationalise the measurement of transfer of learning from modelling using the concepts of close and far transfer, and overconfidence. We investigate learning in discrete-event simulation (DES) projects through an experimental study. Participants were trained to manage queuing problems by varying the degree to which they were involved in building and using a DES model of a hospital emergency department. They were then asked to transfer learning to a set of analogous problems. Findings demonstrate that transfer of learning from a simulation study is difficult, but possible. However, this learning is only accessible when sufficient time is provided for clients to process the structural behaviour of the model. Overconfidence is also an issue when the clients who were involved in model building attempt to transfer their learning without the aid of a new model. Behavioural OR studies that aim to understand learning from modelling can ultimately improve our modelling interactions with clients; helping to ensure the benefits for a longer term; and enabling modelling efforts to become more sustainable.

© 2015 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Going back as far as Churchman and Schainblatt (1965) it has been argued that client involvement in operational research (OR) studies, particularly in model building, is beneficial for successfully implementing a study's findings. Implementation can take the form of concrete changes to the system studied (Kotiadis, Tako, & Vasilakis, 2014), or as learning where the clients gain an understanding that can impact their future decision-making (Robinson, 2014). From this belief that involvement leads to implementation emanates a stream of research on client involvement, for instance, group model building (Andersen, Vennix, Richardson, & Rouwette, 2007; Rouwette, Korzilius, Vennix, & Jacobs, 2011; Rouwette, Vennix, & Mullemkom, 2002) and facilitated modelling (Franco & Montibeller, 2010). In the field of interest in this study, discrete-event simulation (DES), there have been a number of recent papers that focus on how to enhance client involvement in the process of developing and/or using

a model (Adamides & Karacapilidis, 2006; den Hengst, de Vreede, & Maghnouji, 2007; Kotiadis et al., 2014; Robinson, 2001; Robinson, Radnor, Burgess, & Worthington, 2012; Robinson, Worthington, Burgess, & Radnor, 2014). There is, however, very little evidence, other than the anecdotal, that the hypothesised relationship between client involvement and successful implementation actually exists.

In a recent study by the authors, implementation as learning is studied by investigating the so called *high involvement hypothesis*: that client involvement in a DES study improves client learning (Monks, Robinson, & Kotiadis, 2014). We found that involvement in incremental model development and validation, aids clients in the discovery of improvement options or variables previously not considered. The study also provided empirical evidence about the impact of involvement in model building and experimentation on single loop learning (Argyris & Schön, 1996): a change in client attitudes towards management implementation options brought about by involvement in modelling. Although facilitating single-loop learning through models is important the OR literature often comments on the behavioural assumption that modelling within a social context leads to deeper learning about concepts, such as queuing, that can be transferred and adapted to solve future problems.

* Corresponding author. Tel.: +44 023 8120 8201.
  E-mail address: thomas.monks@soton.ac.uk (T. Monks).

In this paper we provide a new analysis and new data to test the impact of client involvement in a DES study on the ability of clients to transfer their learning to another context i.e. a degree of learning more in line with a 'double-loop' (Argyris & Schön, 1996). Here management actions, deep transferrable understanding about a problem and management norms about how to learn are corrected. We contribute new empirical results that involvement in modelling aids learning for future decision making by such a double-loop. We also provide a novel measurement approach for investigating structural and conceptual learning and confidence effects from models. These contributions go beyond our previous work that focused simply on solving an immediate management problem. Our new results and measurement approach focus on the ability of clients to learn more deeply and to recognise the generic lessons that could apply to situations with a similar structure. For example, having realised for a specific queuing problem that resources should not be loaded at 100 percent, the client then transfers that same lesson to a new context involving a similar queuing problem. The benefit of such learning is that the implementation of the DES study extends beyond the immediate problem and it can repeatedly impact upon many problem situations in which the client is involved without the need for further simulation. Behavioural OR studies that aim to understand such learning from modelling can ultimately improve the community of practice's interaction with clients; helping to ensure that benefits are longer term; and enabling modelling efforts to become more sustainable.

The specific objectives of the current study are: to determine if clients can transfer their learning from a DES study to another context; and to determine the effect of client involvement on their ability to transfer that learning. As such, our work aligns with research in behavioural OR that analyses the psychological aspects of model use in problem solving (Hämäläinen, Luoma, & Saarinen, 2013). The problem solving focus means that the study has much in common with research about client learning in group model building in system dynamics (Andersen et al., 2007; Lane, 1994; Rouwette et al., 2002 , 2011; Scott, Cavana, & Cameron, 2014). The main difference is that we use experimental methods to explore the impact of study processes at the individual level.

The remainder of this paper is organised as follows. We firstly present an overview of the theory associated with transfer of learning. This covers both the conditions required for successful transfer and the difficulties people typically face. Secondly, we provide an explanation of the experimental study and materials used in our research. We follow this with the results of the experiment along with a discussion of both the implications and limitations of the work.

2. Transfer of learning

A typical behavioural experiment for transfer of learning consists of analogous training and transfer problems. A participant solves the training problem, is given feedback, and then attempts to solve the transfer problem (Bassok, 2003). An often cited study uses the Tower of Hanoi problem in the training task (Gick & Holyoak, 1980). Participants play the role of a general that must use an army to capture the Tower of Hanoi from a rival general. There are four routes to the tower. If the participant attempts to capture the tower using any single route the army is defeated. However, if the participant divides their army and uses all four routes at once they can overpower the enemy forces and capture the tower. In the transfer problem the participant must decide how to kill a tumour in a patient using X-rays. The problem is that the required dosage of X-ray will damage the tissue it passes through on the way to the tumour. The transfer concept is divide and conquer, i.e. apply lower intensity X-rays from different sides of the body simultaneously.

There are two important factors in the success of the transfer of learning: structural and perceived surface similarity. Structural similarity refers to the underlying mechanics of the problem being stud-
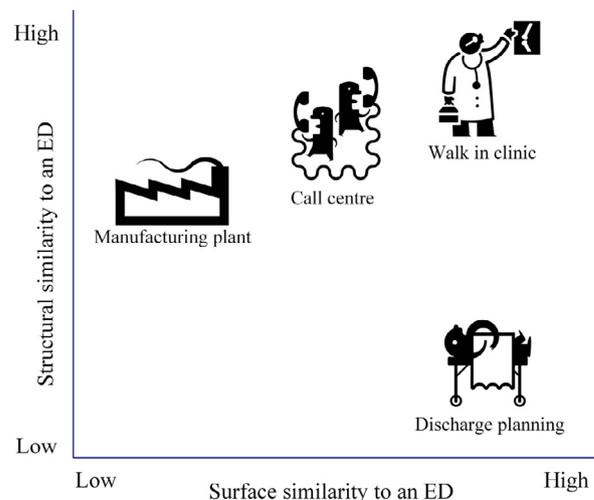

Fig. 1. Dimensions of transfer: surface similarity and structural similarity to a hospital emergency department.

ied. For example, Bakken, Gould, and Kim (1994) study the transfer performance between two system dynamics models with the same underlying feedback structure, but different surface components: a housing market and an oil tanker market. Similarly, in the context of DES both an emergency department and a call centre have an element of structural similarity. They can be considered as queuing systems that are subject to stochastic variation, with much of the variability being driven by arrivals and service times.

Surface similarity refers to whether an individual perceives the training task and transfer task to be similar. Surface similarity is an important cue for initiating a transfer attempt. If an individual does not see any similarity to the training task in the transfer task then they will not attempt to transfer the knowledge. If, however, they do see the similarity, then transfer will be attempted (Bassok, 2003).

Transfer success is most likely when there is both structural and perceived surface similarity. In this case an individual perceives that the new problem is highly similar to one they have tackled before. In addition, the original and new problems are structurally similar so it is valid to transfer the learning from the one situation to the other.

Fig. 1 illustrates the surface and structural similarity of four different queuing problems to a hospital emergency department (the training task). For the healthcare walk in clinic there is a high level of both surface and structural similarity; both concern people in a healthcare setting being seen by healthcare professionals, and both have highly unpredictable inter-arrival times. Structurally a call centre is very similar to an emergency department, but given the different context, the similarity on the surface is not as apparent.

Fig. 1 shows that transfer of learning is probably much less likely from a healthcare context to a manufacturing context. This is, in part, because surface similarity is low; individuals are likely to perceive a large difference between the patients in a healthcare setting and widgets on a production line. Meanwhile, structurally the problems are quite different: in the manufacturing domain variation is driven from cycle times and machine breakdowns as opposed to arrivals and service times.

In a situation where an individual perceives a problem to be highly similar to one they have tackled before, but structurally they are very different, there is a danger of incorrectly transferring learning. One such example is delayed discharges of patients from a hospital. As shown in Fig. 1, this is a problem that might appear very similar on the surface to an emergency department, i.e. there are patients queuing for resource to discharge them. Structurally, however, the discharge of patients is quite different and improving the timeliness of patient discharge involves understanding inter-organisation

co-ordination as well as understanding the trade-off between demand and capacity.

Transfer of learning is classified as either *close* or *far*. The degree of closeness is associated with the perceived surface similarity of the training and transfer problems. For example, a learner may perceive that on the surface a manufacturing domain is very different from a healthcare domain. Successful positive transfer from a healthcare training problem to a manufacturing transfer problem would therefore be classified as *far* transfer. Empirical studies support the view that learning for far transfer is generally more difficult (Barnett & Ceci, 2002). Although this appears to be a fairly straightforward point to grasp it is actually quite difficult to define *far,* as perceived similarity is highly dependent on contextual and individual factors (Barnett & Ceci, 2002). Classification of problems as *far* may refer to different contexts (such as manufacturing and healthcare), the time lag between transfer attempts or the location where transfer is attempted (e.g. a classroom versus a work environment). One method that seems to have a positive effect on transfer is to provide multiple examples of the problem in the training task. Transfer performance appears to improve if both *close* and *far* examples are given during training (Gick & Holyoak, 1980). The explanation appears to be that the differences in surface similarity in the examples improve the individuals' ability to abstract the structure of the problem – giving them a deeper understanding that is transferable.

A final aspect to consider is spontaneous versus informed transfer (Bassok, 2003; Gick & Holyoak, 1987). Consider a manager who is involved in a simulation study of a manufacturing production line. In the study the manager learns that she cannot run production line machines at greater than 90 percent utilisation and still achieve lead time targets. Furthermore, the manager finds that there are huge differences in lead times if machines are run at 85 percent and 95 percent utilisation. Sometime later the manager moves organisation and is put in charge of a new production line. If, unprompted, the manager considers making similar decisions in the new context (i.e. considers lower average utilisation to improve lead times), then the transfer was spontaneous. The manager recognised the similarity between the problems, accessed the relevant knowledge and transferred it successfully.

Informed transfer is where the manager is told to transfer what they have learnt in training. Informed transfer is an artificial procedure only available in a laboratory setting. Its use is in differentiating between problems in accessing relevant knowledge and problems in learning. Many transfer of learning studies employ a design where one group is given a transfer hint (i.e. now use what you have learnt to solve the following problem) and another group is simply presented with the new problem (Bassok, 2003). If the hint group outperforms the no-hint group then it can be concluded that learning is present, but there is an *access problem*, i.e. the no-hint group does not recognise the similarity between the training and transfer problems. If neither group perform well then there would appear to be a learning problem.

In summary, to achieve transfer an individual must not only possess the relevant knowledge, but also perceive the new problem to be similar to a previous problem. Of course, for transfer to be correct the new problem must also be structurally similar to the previous. Perceived similarity can be subjectively divided into close and far; the likelihood of transfer success declining with distance. Improvement of transfer likelihood to far domains comes with increased exposure to analogous problems in different domains. As an analogy consider an experienced DES modeller who has worked in similar projects across manufacturing, healthcare, the public sector and other domains during their career. They are much more likely to think in terms of the structure of the queuing problem than an individual who has had involvement in only a single simulation study. Meanwhile, in daily life, transfer, if it occurs, is spontaneous; there is no need for prompting. However, in the artificial world of the laboratory hints may be needed to encourage subjects to access the relevant knowledge.

3. Transfer experiment

To investigate transfer of learning we sought to design an experiment to explore the degree to which novice simulation users could successfully transfer learning from a recent simulation study to analogous problems in the same and different domains. This investigation involved collecting additional data from the experiments reported in Monks et al. (2014). As such, we do not report the full details of the experiment here, but provide an overview with details of the approach used for measuring transfer of learning.

3.1. Overview of the experiment

The purpose of the experiment is to investigate the hypothesis that greater involvement in model building leads to greater learning about the structure of a problem and so to an improved ability to transfer learning. We created a behavioural experiment making use of novice simulation users (64 business undergraduate students with no simulation training). Bakken et al. (1994) found that MBA students substantially outperformed experts in an experiment that tested for transfer of insights with system dynamics models, since experts tended to refer to their real life experience. As a result, our expectation was that the students would yield a higher rate of transfer success than experts would in our experiment.

The context of the training queuing problem is a fictitious hospital emergency department. To solve the training problem and improve the emergency department performance participants must learn about two concepts:

- There is a non-linear relationship between resource utilisation and the time an entity spends in a process (T1).
- Management (or reduction) of process variation can reduce the time an entity spends in a process (T2).

The simplified process for developing and using the model that we followed with the students allowed the participants to engage in a simulation study in a similar manner as a client would in the real world. For example, participants could question the assumptions in a model, perform face validation checks, request extra detail and define scenarios to be run. At the end of the simulation study we assessed the participants' ability to transfer their learning to eight scenarios containing a description of a queuing problem analogous to the problem faced in the emergency department. We now provide an overview of how the participants were involved in model building and experimentation as well as the three experimental conditions used to vary the degree to which clients are involved in model building.

3.1.1. Involvement in model building

In order to involve participants in model building we created a condensed simulation study process with a modeller and a client. Participants took the role of a client while a researcher provided the modelling expertise. The model begins very simply as a single queue and server model where treatment in ED is represented by a single activity with a single first in first out queue. There are six rounds of refinement to the model where detail is added. The final model is then used for experimentation. Although each participant used the same model for experimentation, the model may evolve quite differently depending on participant choices during the training task. By the end of the model building, all participants have interacted with each of the six models.

Fig. 2 illustrates the procedure used for each refinement of the model. For example, a participant may have chosen to add either doctor resources or split patients into major and minor injuries to the initial single server model. Once this choice had been made a sub-model
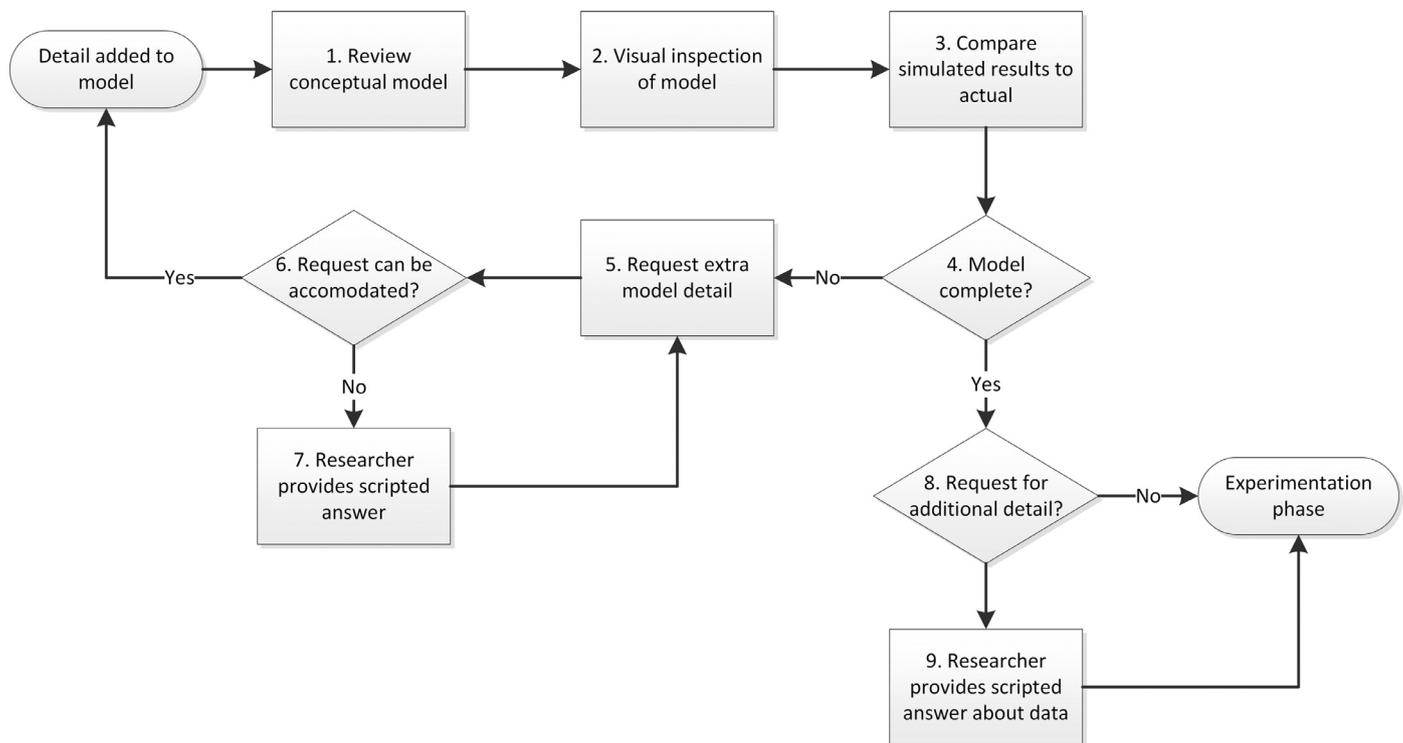
Fig. 2. Model building procedure.

was opened containing the correct level of detail (all combinations are available to the researcher). The participant was presented with a conceptual model summarising the change that has been made, and the remaining simplifications and assumptions. Once reviewed the participant inspected the visual model and could explore a results spreadsheet. The procedure is repeated when the participant asks for another level of detail to be added to the model (e.g. doctor seniority or prioritisation of major injuries). The model building stage ended after the sixth model was completed and all the detail had been added.

Fig. 2 also illustrates that not all participants' requests could be accommodated by the researcher. A common example involves a simplification of doctor multi-tasking (Günal & Pidd, 2006). Emergency department doctors treat multiple patients concurrently and move between them during the patients' stay. In the simulation this concurrent treatment is simplified and modelled as slots, i.e. a doctor can treat four patients concurrently therefore we add four doctor slots for every doctor resource available. If a participant asked for this simplification to be removed the researcher had a scripted answer available: 'it is not possible to remove that simplification as we do not have data available on individual doctor consultations'. In these situations the participant has to reconsider which simplification should be removed. A minority of participants continued to question model simplifications (or requested additional model detail) once the model was 'complete'. This is not unlike standard client face validation processes in real studies. In the instances where additional detail was requested the scripted answer was that the study did not have sufficient data to include this detail. In the instance where the request related to an assumption, the scripted answer was to advise the participant that assumptions could be explored during the experimentation phase.

3.1.2. Involvement in experimentation

For the experimentation participants again acted as the clients. Participants directed the researcher with regards to which scenarios were run while the researcher provided guidance on what is possible. Participants were able to watch the model running in visual mode

Table 1
Experimental conditions.

Condition	Time (hours)	N	Description
MR	1.25	22	Model reuse. Participants reuse a pre-existing model to investigate the emergency department problem. There are six predefined scenarios. Participants have free choice over remaining scenarios.
MBL	1.25	21	Model building with limited experimentation time. Participants are involved in building the model. They only have 15 minutes for experimentation, enabling them to explore three scenarios.
MB	2	21	Model building. Participants are involved in building the model. They have a further hour to experiment with the model. Same experimentation conditions as MR.

and review batch run results. A spreadsheet tracked all scenarios run and provided a summary so that the participants could review their experimentation history.

3.1.3. Three types of client involvement

We manipulated participant involvement in model building and experimentation by including three experimental conditions which are summarised in Table 1. Low involvement was investigated in a condition where participants reuse a pre-existing model ('model reuse' – MR). High involvement was investigated in two conditions. In both the participants were involved in the same model building process. However, in the first high involvement condition the budget of time available for building and using the model is equal to that for MR ('model building with limited experimentation time' – MBL). As a result they only had limited time for experimentation; enough to explore three scenarios. In the second high involvement condition ('model building' – MB), participants were given

Table 2
Description of transfer problems.

Transfer	Problem	Context	Scenario	Transfer cues	Concept
Close	S1	GP surgery waiting times	Participants must consider two options: running the clinic as a drop-in system or making use of appointment slots.	• Queues/waiting times • Variable IAT and process times (people)	T1
	S2	Emergency department waiting times	Participants must consider how to manage the resource levels on a shift in the emergency department in face of an increase in demand. Summary statistics for the waiting time and utilisation of resources are provided.	• Queues/waiting times • Variable IAT and process times (people) • Resource utilisation statistics/shifts patterns	T2
	S3	Operating theatre waiting times	Participants are asked to consider three options for reducing theatre waiting times. Attempt to maximise current utilisation, increasing capacity and extending the duration of slots.	• Queues/waiting times • Variable process times (people) • Resource utilisation statistics	T2
	S4	Receptionist waiting times	A walk in centre and a GPs surgery are being co-located. Participants are asked to consider if the queues for reception should be pooled or un-pooled.	• Queues/waiting times • Variable IAT and process times (calls/people) • Queue pooling	T1
Far	S5	Food manufacturing cycle time	Participants are asked to how they would reduce the lead time of a three part assembly process with deterministic arrivals and no breakdowns. Summary statistics of the utilisation of each work station are provided.	• Queues and lead time • Resource utilisation statistics • Variable process times (food widgets)	T2
	S6	Police call centre waiting times	Participants must consider if performance will be improved at a call centre, by increasing resource, pooling resource or creating dedicated resource.	• Queues/waiting times • Variable IAT and process times (calls/people) • Queue pooling	T1
	S7	Food manufacturing cycle time	Participants are asked how they would maintain cycle time of a four part assembly process with a bound queue. Summary statistics are provided for process and machine breakdown times.	• Queues/lead time • Variable process times (food widgets) • Inter-arrival times (breakdowns)	T1
	S8	Service call centre waiting times	Participants must consider how manage the resource across four shifts at a service call centre. There are options for hiring additional call takers, reallocation of resource and splitting shifts. Summary statistics are provided for the utilisation and performance of each shift.	• Queues/waiting times • Variable IAT and process times (calls/people) • Resource utilisation statistics/shift patterns	T2

45 minutes longer enabling them to carry out much more extensive experimentation. This means they had equivalent time for experimentation as the MR participants. For the experiment the students were randomly assigned to the three conditions to give 22 students in the MR group, 21 in the MBL group and 21 students in the MB group. We make use of randomisation in the allocation of participants to groups to mitigate the risk of imbalances across groups in factors such as intelligence and experience of problem solving.

Participants in the MR group were introduced to simulation in the same manner as the MB and MBL groups, as detailed in Section 3.1.1. However, there were no iterative steps in building the model. Instead, once the introduction to the simulation was complete, MR participants were presented with the complete and final model as used by the participants in MB and MBL. The MR participants were then talked through the logic of the model using a script. Participants in all conditions were given the same documentation about the model.

3.2. Measuring transfer of learning

We created eight transfer problems with varying distances of surface similarity to the emergency department problem. All participants answered these in the same order. For simplicity we classify these as either close or far transfer problems. The four close transfer problems are set in a healthcare context. These scenarios have high surface similarity and high structural similarity to the training problem. In contrast the far transfer problems have low surface similarity to the training problem, but are still structurally similar (i.e. a queuing problem). The far transfer problems are either set in call centres or a food manufacturing plant. Each problem details a scenario, provides transfer cues, lists multiple choice answers and provides space for a qualitative answer. We include the qualitative answer in order to accurately separate the random chance of selecting a correct answer from the successful transfer of learning. Transfer cues are pieces of information that prompt participants to recall the case study problem. For example, queues of customers in a scenario should prompt participants to recall that the ED problem was based around queue management. Each scenario was designed to test for one of the two transfer concepts listed in Section 3.1.

The eight transfer scenarios, and the associated transfer cues and concepts, are summarised in Table 2. We presented these to each participant using a questionnaire following the training task. The training problem contains numerous examples of the two transfer concepts in action. For example, participants ran experiments analysing the allocation of nurses to shifts and the use of resources to learn

about resource utilisation and performance. Scenarios S2 and S8 listed in Table 2 test for transfer of this learning (T1). Scenario S2 presents the problem of resource utilisation at another similar hospital where performance is currently good, but demand is about to increase. To transfer the learning successfully participants must correctly reason that the loss of spare (buffer) capacity will likely reduce average performance against targets. Far transfer is tested by Scenario S8 set in a service call centre. For successful transfer of learning the participants must identify that the better performing shifts have more spare capacity.

An example of concept T2 in the training problem is the management of the variation of different kinds of patient emergencies by pooling treatment cubicles. Scenarios S4 and S6 test for transfer of this concept. Scenario S4 asks participants to decide how waiting time could be minimised when checking into a co-located walk in centre and GP surgery. For successful transfer participants must correctly reason that combining the queues is a more efficient way to manage the variation in types of patient arriving. Far transfer is tested by scenario S6, a police call centre context, where participants are asked how to reduce caller waiting time to a number of small regional call centres. For successful transfer of learning the participants must recognise that a single larger call centre manages the variation in regional demand more efficiently. Full details of the scenarios can be found in the online supplementary material.

3.2.1. Dependent variables

We measured two types of dependent variables: success in transfer of learning to problems analogous to the training problem and the overconfidence in the participants' answers. There are three transfer of learning measures: total, close and far transfer. Overconfidence has two measures, overestimation of ability to correctly transfer and the proportion of errors made in high confidence.

3.2.2. Questionnaire

Transfer of learning was assessed using a post-test questionnaire consisting of eight scenarios that were developed over a pilot of 16 participants. Each transfer scenario asked participants to select a multiple choice answer, provide a qualitative answer describing their reasoning and report the confidence they had in their answer. All participants were given the same scenarios in the same order and provided the same information

Measurement of the confidence participants had in their answers to the transfer scenarios is based on scales used to measure metacognitive confidence (Petty, Briñol, & Tormala, 2002). After answering each transfer problem participants were asked to rate the confidence they have in their answer on a nine-point scale (1 = not at all confident to 9 = extremely confident). Participants were told that answers of above five reflected the belief that their answer was correct. Post-hoc analysis revealed that only 18 percent of answers were below five indicating that the vast majority of participants believed they were successful in solving each transfer problem.

As an example of a transfer scenario, consider call centre scenario S6 described in Table 3. This scenario provides three options to choose from to reduce call centre waiting times. Participant MR1 believed that waiting time performance would be improved by choosing the option that split the call centres into smaller geographic regions. Participant MR1 provided the following qualitative answer:

"[Option A allows] specialisation of tasks increasing focus and speed. Obviously if [the call centres are] still not meeting targets [then] increase staffing and lines. Combining everything (option c) complicates tasks and therefore increases queues"

Participant MR1 was very confident that this answer was correct and scored herself as an eight out of nine.

Table 3
An example scenario: call centre scenario S6.

Scenario	The waiting time of a number of regional emergency call centres needs to be improved.
Options	a. Split call centres into smaller geographic regions (more queues).
	b. Hire more call takers (more resource).
	c. Create a single large call centre to take all calls (less queues).
Transfer success	Transfer is achieved by recognising that performance is driven by the variation in inter-arrival time variability. A combined call centre – answer c – manages the peaks and troughs in inter-arrival variation more efficiently.

3.2.3. Scoring transfer of learning

The answers were coded twice by one of the authors (Monks) with a six month time gap. To minimise any bias the participant details were hidden from view during coding and the order of the answers was randomised in each coding. The coding had a high reliability, $\alpha = 0.87$, and it had a high intra-class correlation coefficient, $0.838 < r < 0.895$ (91 percent of the codings were the same). The differences between each coding typically reflected cases where the participants had provided minimal detail on their reasoning. The differences were resolved by reviewing the comments made in each coding and agreeing a final score.

As an example of the coding procedure consider the information cues provided to a participant by scenario S2 detailed in Table 2. The performance target and percentages provide the cues to recall the performance of the simulation model. The participants were asked if they agree that more staff should be introduced. All participants were asked to give the reasons why they think their choice would improve the performance of the system. An example answer provided by a participant is:

"As the workload is going to increase, it is important to increase the number of staff so that they can cope. Depending on the increase in workload it may also be necessary to increase the number of cubicles, but is unlikely as their utilisation is currently lower than staff utilisation and won't reach its maximum"

The second sentence provides the most detail on the participants' reasoning. The participant was thinking in terms of maximum capacity. If the resources are working at less than 100 percent then they can cope. If 100 percent is exceeded then more resource is necessary. Whilst this seems sensible it fails to transfer any learning from the training problem in which it was identified that even with utilisation below 100 percent, resources cannot cope when there is variability in arrival and service times. An example of successful transfer would discuss performance levels and the relationship to utilisation: even if utilisation is increased to 90 percent there will be a change in the performance of the system. Hence the example answer above is coded as a zero – failure to transfer learning.

Subjectivity in the coding procedure was minimised by constructing focused transfer scenarios testing a single transfer concept and by issuing a standardised model answer to the coder. As importantly, the pilot experiments provided insight into the phrasing of participant responses and the associated understanding of the transfer concept. We illustrate the potential differences in participant answers for transfer scenario S6 in Table 4. This lists the participant identifier, their qualitative answer, if the correct multiple choice answer is chosen and if the reasoning has been coded as correct. The first three examples are coded as successful transfer and illustrate the range of phrasings that participants used to describe the benefits of queue pooling to handle variability. Although different phrasings are used it is clear that each participant recognises variability in inter-arrival

Table 4
Example qualitative answers to transfer scenario s6.

Participant	Answer	Correct answer	Correct reasoning
MB12	"[This] makes [the] police call centres more flexible. If demand in one geographic area is high, additional call takers, who [were] working in different areas previously, are available "	1	1
MBL11	"Put everyone in one queue. This means that [there is] no area operators overloaded whist another area has free [operators]. It levels supply."	1	1
MR19	"It's all about resource utilisation, combining all operations into a single call centre would be rational. Idle call takers would be able to help other call takers who have queues …."	1	1
MB7	"… transferring information between various stages [of the call centres] consumes time. A single call centre would increase efficiency and reduce it."	1	0
MBL4	Emergency demand in [the] different regions will be uniform, so only way is to increase staff	0	0

rates and service times lies at the heart of the problem. Incorrect qualitative answers often stood in stark contrast to successful transfer. MB7's answer in Table 4 illustrates this point i.e. the participant argued that a larger call centre is more efficient (lower average service times) than multiple smaller call centres. The final participant MBL4 in Table 4 chose an incorrect answer (more call operators), but again illustrates the clear cut nature of incorrect answers. Here the participant argues that the demand in each region is 'uniform' (meaning constant and equal) and hence there is no benefit in pooling.

3.2.4. Scoring overconfidence

Rather than simply assessing incorrect transfer we adopted two measures of overconfidence in transfer. This allowed us to distinguish between an erroneous answer made because the participant was unsure what to do and erroneous answers that would potentially be acted upon by the participant. The two overconfidence measures are constructed from the participant's confidence scores on each transfer scenario. Overestimation of their own abilities (Bendoly, Croson, Goncalves, & Schultz, 2010) is the difference between the proportion of problems to which a participant predicts that they have successfully transferred learning and the actual proportion where learning is successfully transferred. For example, if a participant predicts that they have correctly answered six out of the eight problems correctly (75 percent), but have actually only answered four correctly (50 percent) then they have overestimated their ability by 25 percent. Secondly, overconfidence was measured as the proportion of errors that are made with high confidence. High confidence is classed as a reported confidence score (Section 3.4.2) of greater than five. Therefore, a high confidence error is an incorrect answer made with a confidence score of greater than five. As the choice of the cut-off for high confidence errors was a judgement (i.e. the upper half of the scale), we include a sensitivity analysis of results using a more stringent cut-off of eight out of nine (section 4.3).

4. Results

The results from the experiment are now presented firstly with respect to transfer success for each scenario, then the transfer performance of the experimental conditions, and finally the results for the confidence of the participants are presented.

4.1. Transfer performance by scenario

Overall transfer of learning successfully occurred most frequently in the healthcare scenarios (44 percent), followed by call centre scenarios (35 percent) and least frequently in the manufacturing scenarios (13 percent). This was as expected given the levels of surface similarity. Transfer of learning across the two transfer concepts was similar ($T1 = T2 = 34$ percent).

Fig. 3 presents transfer of learning success and the associated confidence for each scenario ordered by transfer success. This indicates the level of surface similarity that the participants perceived between the training task and each scenario. Participants generally performed better in those scenarios that we initially classified as close transfer (S1 to S4) than those we classified as far transfer (S5 to S8). An exception is scenario S8 that is ranked as the second most successful scenario and a substantial improvement over the other far transfer scenarios.

Scenario S8 is set in a call centre and tests the transfer of participants learning about resource utilisation and performance. It is similar to the training task as there are four staff shifts across the day. Participants see the performance (percentage of calls answered in a target time) of each shift individually along with the utilisation of staff. Transfer success is achieved by recognising that underperforming shifts are so busy (high utilisation) that they are into the non-linear explosion of customers waiting time (concept T1). In addition any increase in staff resource utilisation on the shifts currently meeting targets will reduce their performance against the call answering target. The majority of participants (55 percent) successfully transferred their learning from the emergency department training task to this scenario.

An explanation for the participants recognising the structural similarity of scenario S8 is that this scenario may be closer in surface similarity to the training task than expected. In particular, the emergency department model has specific examples about shift reallocation. One participant from the MBL group even referred to the behaviour of the simulation model outputs in his answers. In the remaining analysis we reclassify close transfer to incorporate scenarios S1–4 and S8.

4.2. Transfer performance by experimental condition

As a reminder we measured transfer of learning at three levels: total, close and far transfer. Transfer of learning results by condition are reported in Table 5. Total transfer was significantly different across groups with MB participants typically transferring learning to one problem more than MBL and MR. Overall MR and MBL participants performed similarly. The majority of the difference in total transfer between groups is explained by our revised classification of close transfer performance (S1–4, S8); MB outperformed MBL and MR. All groups performed poorly in the far transfer scenarios on average answering one or less correctly.

Fig. 4 presents the transfer of learning results by scenario and group. Scenarios are presented in the order of perceived surface similarity observed across all participants. There are two notable points. Firstly performance across scenario S1, which has the most surface similarity, is similar across all conditions. Secondly, participants in the MB group outperformed the MBL and MR participants across scenarios S3, S4 and S5. The MB group also performs at a similar level across scenarios S2, S3, S4 and S8 (range = 43–52 percent). The slightly lower performance of MB in S8 is not significant ($X^2(2) = 0.700$, $p = 0.731$, Kruskal–Wallis).

Performance across the far transfer scenarios (S5–7) is consistently low across groups and small differences cannot be explained by anything other than chance. The largest difference exists in scenario S5, set in a manufacturing plant, where 29 percent of MB participants could correctly identify that the process bottleneck was a
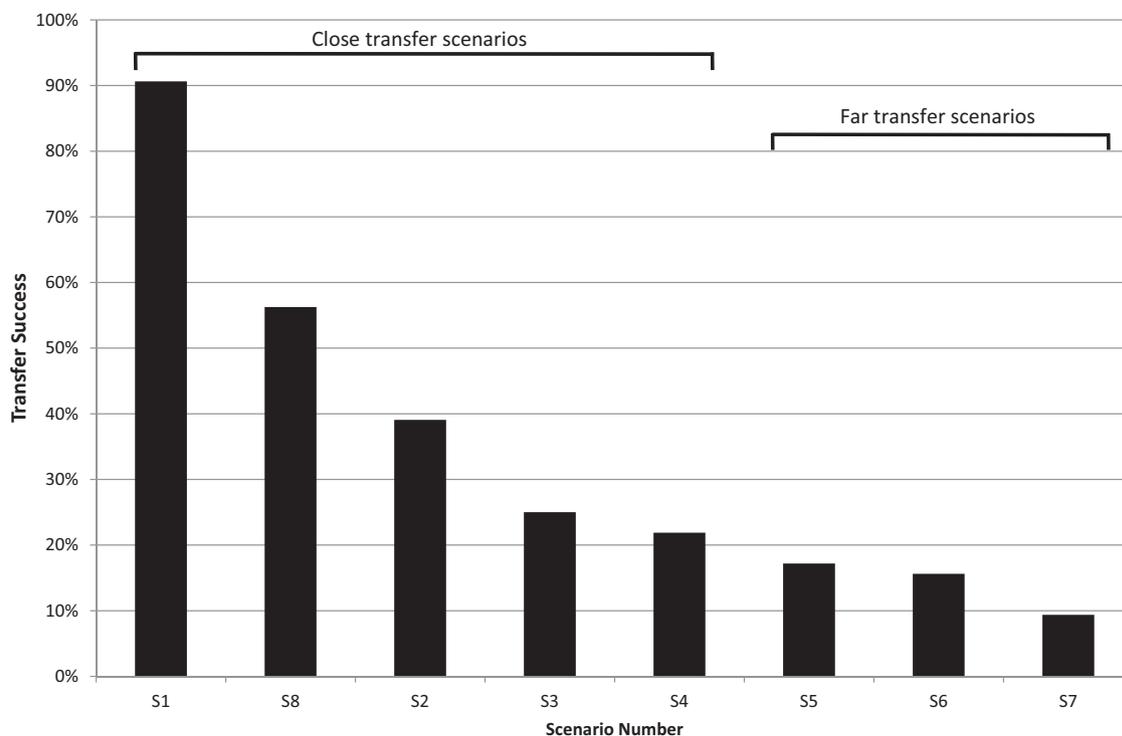
Fig. 3. Perceived similarity of transfer scenarios to the emergency department training.

Table 5
Transfer of learning results.

	Mean (St. Dev.)			Group differences	Mean difference (95 percent CI); *p*-value†		
	MB	MBL	MR		MB–MR	MB–MBL	MBL–MR
Total transfer	3.4 (1.4)	2.3 (1.1)	2.5 (1.1)	F = 4.53; p = 0.015**	0.9 (0.1–1.6); p = 0.008***	1.0 (0.3–1.8); p = 0.035**	−0.2 (−0.9–0.5); p = 0.533
Close transfer	2.8 (1.1)	2.0 (0.9)	2.1 (1.0)	F = 7.32; p = 0.028**	0.7 (0.1–1.3); p = 0.039**	0.8 (0.2–1.4); p = 0.016**	−0.0 (−0.5–0.6); p = 0.756
Far transfer	0.6 (0.7)	0.3 (0.5)	0.4 (0.6)	F = 0.86; p = 0.289			

Close transfer scenarios: S1–4, S8. Transfer measured as number of correct answers coded with correct reasoning.
† Significance key: **$p < 0.05$; ***$p < 0.01$; F = one way ANOVA.

highly utilised work station compared to 14 percent in MBL and 9 percent in MR.

Overall confidence in the participants' answers was high and consistent (median = 7.0; inter-quartile range = 2.0) with no significant difference between the groups (Kruskal–Wallis: $H(2) = 2.7$, $p = 0.252$). Fig. 4 illustrates only minor variation in confidence across conditions.

4.3. Overconfidence

We measured overconfidence in terms of participant's overestimation of their own abilities and the proportion of errors committed in high confidence. Full test results and summary statistics are reported in Table 6 . Participants in all three groups overestimated their performance by an average of 47 percent (95 percent CI 40.3–53.1 percent). There was evidence that this was different across groups. Participants in the MBL ($mean_{MBL} = 60$ percent) group overestimated their performance more frequently than MB participants ($mean_{MB} = 45$ percent). However, the result for the comparison of MBL and MR was not significant at the 95 percent level ($mean_{MR} = 47$ percent; mean difference = 12.7; 90 percent CI 0.9–24.5). There was no evidence of a difference in overestimation between MR and MB.

Our second measure of overconfidence, the proportion of errors committed in high confidence, was different across groups. Participants in the MBL group typically made a higher proportion of their errors in high confidence than MR. The weaker evidence of a higher

proportion in MB than MR (mean difference = 12.8; 90 percent CI 1.7–24.0) was validated by bootstrapping (95 percent CI_{Bca} 0.2–25.0, $p = 0.047$; replications = 2000).

As the choice of the cut-off for high confidence errors was a judgement, we conducted a sensitivity analysis of results using a more stringent cut-off of eight out of nine (112 answers were rated eight or above). Again participants in the MBL group made a higher proportion of their errors with high confidence than MR (mean difference = 12.0; 95 percent CI 1.0–23.0). The results for the MB versus MR were not significant at either the 90 percent of 95 percent level using standard tests (mean difference = 8.9 percent, 95 percent CI −2.6–20.4).

5. Discussion

We now discuss the results with respect to the research objectives: to determine if clients can transfer their learning from a DES study to another context; and to determine the impact of client involvement on their ability to transfer that learning. We also discuss the impact of involvement on confidence, the limitations of the study and the contribution of the work to behavioural OR.

5.1. Can DES clients transfer learning?

Transfer of learning to the analogous problems proved difficult for our participants on average only achieving transfer success for
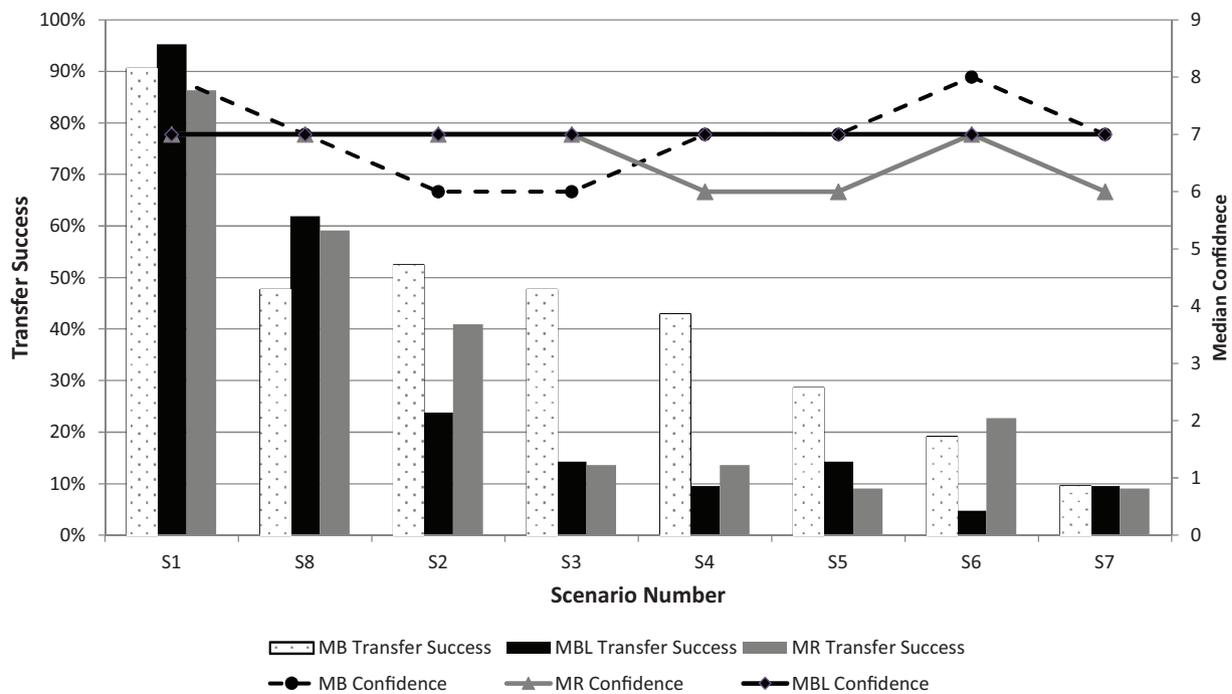
Fig. 4. Transfer performance and median confidence in answers by scenario and group.

Table 6
Overconfidence results.

	Mean (St. Dev.)			Group differences[†]	Mean difference (95 percent CI); *p*-value[††]		
	MB	MBL	MR		MB–MR	MB–MBL	MBL–MR
Overestimation percent	45 (18)	60 (29)	47 (23)	$F = 2.87$; $p = 0.06$*	2.4 (−10.8–15.6); $p = 0.713$	15.1 (2.1–28.1); $p = 0.023$**	12.7 (−1.5–27.0); $p = 0.078$*
High confidence errors percent	49	56	37	$\chi^2 = 9.50$; $p = 0.009$***	12.8 (−0.3–26); $p = 0.058$*	6.8 (−6.5–20); $p = 0.319$	19.6 (7.2–32.0); $p = 0.002$**

[†] F = one way ANOVA; χ^2 = Chi-squared test of association.
[††] Significance key: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

three out of eight scenarios. The majority of this transfer success was achieved in the scenarios we pre-specified as close transfer. These scenarios had high surface and structural similarity to the emergency department training problem i.e. a healthcare context where it is people who queue. A large group of participants were also able to successfully transfer learning to a scenario with lower surface similarity to the emergency department. The result for scenario S8, a call centre context, demonstrates that participants identified a structural similarity between the transfer and the training problems. In fact, performance was so high in scenario S8 that it is ranked higher than scenario S2 set in an emergency department. We defined surface similarity in terms of application domain (for example, a healthcare scenario bares more resemblance to the training task than a call centre scenario). These results suggest a more subtle effect. Both scenarios consider shifts of workers, resource utilisation and variable inter-arrival times. However, a key difference is that scenario S2 considers only one shift while, in the same manner as the emergency department training problem, scenario S8 considers four shifts of workers. This subtle difference appears to have given scenario S8 more surface similarity to the training problem and led to access problems in scenario S2.

In summary, we demonstrate that, in line with theory, transfer of learning from a simulation study to another problem is difficult with subtle reasons why one problem may have more surface similarity than another. However, given some limited success in scenario S8, a

scenario we originally classed as far transfer, we are able to provide evidence that involvement in simulation studies does enable deep structural understanding of the problem studied. This brings us to our second research objective exploring the effect of three different types of study involvement on learning.

5.2. The effect of involvement on transfer

Out of the three groups, participants in MB scored more highly than the two groups with a tighter time constraint. An explanation for this difference then appears to be that more time was needed to process the complexity of the transfer concepts and hence increase transfer success; time provided in the MB group. However, it is clear that MBL and MR still learnt about the structure of the queuing problem in the emergency department, as a majority could transfer this learning to scenario S8. One explanation therefore is that MBL and MR had learnt the relevant knowledge, but could not access it as easily as the MB group. This has implications for the practice of DES. In particular involving clients is not a panacea and some thought should be given to the length of the experimentation phase of the project. This correlates with evidence from experimentation practice that shows that more experienced modellers have slightly shorter model building phases, allowing more time for experimentation (Hoad, Monks, & O'Brien, 2015).

A related question to our findings is how DES can be most effectively used to support training of managers. Learning from simulation has often been studied from a serious gaming perspective (Lane, 1995; Langley & Morecroft, 2004; Rouwette, Größler, & Vennix, 2004; van der Zee, Holkenborg, & Robinson, 2012; van der Zee & Slomp, 2009) with mixed reports on its success as a learning aid (Lane, 1995). Our results suggest that *access to learning* from serious gaming and the *transfer* of that learning is improved if serious games are designed so that managers are allocated additional time for involvement in model building, or at least the conceptualisation of the model. Meanwhile, this should not come at the expense of the time available for experimentation (gaming) with the model. Our methodology to measure transfer of learning is also highly applicable to assess the benefits of training with DES and could be incorporated in to the design of serious games.

5.3. The effect of involvement on overconfidence

The previous section highlighted that it is the *time available* for client involvement in both model building and experimentation that is associated with transfer success and access. For our secondary measure – overconfidence – we can conclude that the degree of client involvement in model building has an effect.

In general the confidence of participants was consistent and high across the transfer problems. Participants overestimated (over predicted) their own performance on average 47 percent of the time. There was evidence that MBL group has the highest overconfidence; overestimating their score on average 60 percent of the time compared to 45 percent and 47 percent in MB and MR, respectively. There was also strong evidence that when an error was committed that it was typically with higher confidence in participants that had been involved in model building (MB make less mistakes overall, but when they do make a mistake it was with higher confidence than MR).

We propose two mechanisms at work that affect the difference in the overconfidence of participants. Firstly, involvement in model building may affect what we term as *ownership pride*. That is, involvement in building the model biased and inflated a participant's beliefs about their own knowledge. When combined with difficulties in accessing knowledge in MBL this led to increased overestimation. Secondly, we cannot rule out that the model reuse participants suffered to some extent from *not invented here syndrome* (Pidd, 2002; Robinson et al., 2004). That is, trust in the model is more difficult when you have not had some involvement in its development. Difficulties in trust would perhaps prevent the inflation of a participant's belief in their own knowledge. To investigate these propositions, further research could replicate our experiment and add a group that does not use simulation during training. There are two methodological issues to overcome in such a design. Firstly, what feedback should be given to participants in the training task and how should it be delivered? Secondly, what level of experience in simulation should participants have: none or some degree of training?

5.4. Limitations

A key factor effecting transfer is the background knowledge of the decision maker (Gick & Holyoak, 1987). As we used students, the participants in the experiment lacked familiarity with the system. Thus a proportion of their cognitive processing power was put to work on building familiarity with the system while building understanding about DES in general and emergency department performance. Participants may also expend effort attempting to work out the experimental hypothesis (Field & Hole, 2003); clearly this is not the case in real world studies.

To an extent this increased mental effort is compensated for by the simplified training problem. However, decision makers within a real system may find it easier to concentrate on understanding the problem and hence improve transfer likelihood. We also note that the opposite can happen: increased background knowledge of decision makers can lead to transfer difficulties. In an experiment testing for transfer of insights between analogous System Dynamics models, researchers found that students substantially outperformed experts (Bakken et al., 1994). Experts would 'take the actions they would have taken in real life' (Bakken et al., 1994) rather than experiment with different approaches to see the effect.

An alternative explanation of the lower MBL transfer results could be that the complexity of the model was too low for the benefits of involvement in model building to take effect. In a large complex process, for example of a whole hospital (Günal & Pidd, 2011), supply chain (Katsaliaki & Brailsford, 2006) or whole healthcare system (Brailsford, Lattimer, Tarnaras, & Turnbull, 2004), decision makers may be more likely to think in terms of local optimisation than our participants. The transferrable learning therefore might be an individual's approach to problem solving rather than the queuing concepts we tested. Framing the problem in this manner would mean that successful transfer would involve recognising that (possibly delayed) effects of changes may occur downstream in a process.

Although we controlled the models that participants could use in the experiment, we provided participants the freedom to build the model in an order they chose. Further work employing a similar methodology should consider exploring if such an approach has any practical effect on transfer success.

Finally the overconfidence effects that we saw in the model building groups may be associated with the population of participants in the experiment. In particular, students who are more familiar with assessments may be more confident in their answers than managers working in industry. Although we again note that the overconfidence effect was systematically different between the model building and reuse groups.

5.5. Contribution to behavioural OR

Our study provides a novel measurement approach for behavioural OR studies that aims to analyse the impact of modelling and model use in long term problem solving and decision making. In particular, our approach is the first to operationalise the measurement of transfer of learning from modelling using the concepts of close and far transfer, and overconfidence. A strength of this approach is that it has the ability to highlight access issues (i.e. the inability to access learned knowledge) given particular types of 'training' and hypothesise learning mechanisms that clients experience when involved in model building.

Our empirical findings add to the limited, but growing, evidence base that is beginning to support the belief that client involvement in modelling facilitates learning. It also tests the high level behavioural assumption that modelling leads to deep conceptual learning about a problem that can be transferred to future problems.

We grounded our approach in a psychological learning framework, transfer of learning theory, (Bassok et al., 2003) and to aid further behavioural OR research in this area we provide a detailed review and introduction to this area. Given these contributions, our multiple close and far scenario approach is highly adaptable to DES application domains other than healthcare. Further observational studies might consider applying the approach to real clients following a simulation study, although this will rely on researchers having access to suitable material to draw on. Behavioural factors relating to clients and model use are often associated with studies in problem structuring methods (e.g. Franco, 2013; Franco & Rouwette, 2011; Franco, Meadows, & Armstrong, 2013) or system dynamics (e.g. Hovmand et al., 2012; Korzilius, Raaijmakers, Rouwette, & Vennix, 2014; Rouwette et al., 2011; Rouwette, Bleijenbergh, & Vennix, 2014; Shields, 2001, 2002). We extend this literature to the study of model development and use in DES which is one of the most prominent hard OR

methods used in practice (Jahangirian, Eldabi, Naseer, Stergioulas, & Young, 2010). Our focus on clients' use of models complements other behavioural DES research, such as Onggo and Hill (2014), that focuses on how modellers conduct their simulation studies.

6. Conclusions

Practical simulation studies that aim to help clients and businesses better manage their processes should take note our results. It is often assumed that involvement of clients in modelling leads to (*double-loop*) learning about the underlying structure of processes and systems under study. Our results demonstrate that, in line with learning theory, such learning is difficult to transfer to analogous problems even in the same domain. We found transfer of learning from a simulation study is difficult, but possible. The main difficulty for participants was recognising the structural similarity of the transfer and training problems; however, access problems were reduced when participants were provided sufficient time for involvement in both model building and experimentation. In the real world, where the surface similarity of analogous problems is even more obscured and budget is tight, we should expect this to be even more difficult for clients. Maybe the primary learning that can be transferred from a DES study is that the problem addressed was non-trivial and that when a similar problem is encountered in the future it would be useful to employ DES again.

A novel contribution of our behavioural research is that the level of involvement in model building and experimentation affects overconfidence in subsequent decisions; specifically we saw that model building groups made a higher proportion of high confidence errors than model reusers. Further fieldwork research is needed to replicate and understand these confidence effects. For instance are differences due to model *ownership and pride*, achieved through involvement in model building, or due to *not invented here syndrome*, an issue with model reuse.

As the field of behavioural OR emerges, the body of experimental and fieldwork studies that analyse modelling and model use offer a unique opportunity to improve the knowledge of how to run modelling projects to achieve maximum benefit both in terms of improved understanding and improved likelihood of implementation of findings. Our study provides results and a novel measurement approach to further these aspirations.

Acknowledgements

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2015.08.037.

References

Adamides, E. D., & Karacapilidis, N. (2006). A knowledge centred framework for collaborative business process modelling. *Business Process Management Journal, 12*(5), 557–575. doi:10.1108/14637150610690993.

Andersen, D. F., Vennix, J. A. M., Richardson, G. P., & Rouwette, E. A. J. A. (2007). Group model building: Problem structuring, policy simulation and decision support. *The Journal of the Operational Research Society, 58*(5), 691–694. doi:10.2307/4622747.

Argyris, C., & Schön, D. A. (1996). *Organisational learning II: Theory, method and practice*. Addison Wesley.

Bakken, B., Gould, J., & Kim, D. (1994). Experimentation in learning organisations: A management flight simulator approach. In J. Moorcroft (Ed.), *Modelling for learning organisations*. Portland: Productivity Press.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637.

Bassok, M. (2003). Analogical transfer in problem solving. In J. E. Davidson, & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 343–369). Cambridge: Cambridge University Press.

Bendoly, E., Croson, R., Goncalves, P., & Schultz, K. (2010). Bodies of knowledge for research in behavioral operations. *Production and Operations Management, 19*(4), 434–452. doi:10.1111/j.1937-5956.2009.01108.x.

Brailsford, S. C., Lattimer, V. A., Tarnaras, P., & Turnbull, J. C. (2004). Emergency and on-demand health care: Modelling a large complex system. *Journal of the Operational Research Society, 55*(1), 34–42.

Churchman, C. W., & Schainblatt, A. H. (1965). The researcher and the manager: A dialectic of implementation. *Management Science, 11*(4), 69–87. doi:10.2307/2628012.

den Hengst, M., de Vreede, G. J., & Maghnouji, R. (2007). Using soft OR principles for collaborative simulation: A case study in the Dutch airline industry. *Journal of the Operational Research Society, 58*(5), 669–682. doi:10.1057/palgrave.jors.2602353.

Field, A., & Hole, G. J. (2003). *How to design and report experiments*. London: Sage Publications Ltd.

Franco, L. A. (2013). Rethinking Soft OR interventions: Models as boundary objects. *European Journal of Operational Research, 231*(3), 720–733 http://dx.doi.org/10.1016/j.ejor.2013.06.033 .

Franco, L. A., Meadows, M., & Armstrong, S. J. (2013). Exploring individual differences in scenario planning workshops: A cognitive style framework. *Technological Forecasting and Social Change, 80*(4), 723–734 http://dx.doi.org/10.1016/j.techfore.2012.02.008 .

Franco, L. A., & Montibeller, G. (2010). Facilitated modelling in operational research. *European Journal of Operational Research, 205*, 489–500. doi:10.1016/j.ejor.2009.09.030.

Franco, L. A., & Rouwette, E. A. J. A. (2011). Decision development in facilitated modelling workshops. *European Journal of Operational Research, 212*(1), 164–178 http://dx.doi.org/10.1016/j.ejor.2011.01.039 .

Gick, M. L., & Holyoak, K. J. (1987). The cognitive basis of knowledge transfer. In S. M. Cormier, & J. D. Hagman (Eds.), *Transfer of learning: Contemporary research and applications* (pp. 9–46). London: Academic Press.

Gick, M. L., & Holyoak, S. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*(3), 306–355. doi:10.1016/0010-0285(80)90013-4.

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

Günal, M., & Pidd, M. (2011). DGHPSIM: generic simulation of hospital performance. *Transactions on Modeling and Simulation (TOMACS), 21*(4). doi:10.1145/2000494.2000496.

Hämäläinen, R. P., Luoma, J., & Saarinen, E. (2013). On the importance of behavioral operational research: The case of understanding and communicating about dynamic systems. *European Journal of Operational Research, 228*(3), 623–624 http://dx.doi.org/10.1016/j.ejor.2013.02.001 .

Hoad, K., Monks, T., & O'Brien, F. (2015). The use of search experimentation in discrete-event simulation practice. *Journal of the Operational Research Society, 66*(7), 1155–1168.

Hovmand, P. S., Andersen, D. F., Rouwette, E., Richardson, G. P., Rux, K., & Calhoun, A. (2012). Group model-building 'scripts' as a collaborative planning tool. *Systems Research and Behavioral Science, 29*(2), 179–193. doi:10.1002/sres.2105.

Jahangirian, M., Eldabi, T., Naseer, A., Stergioulas, L. K., & Young, T. (2010). Simulation in manufacturing and business: A review. *European Journal of Operational Research, 203*(1), 1–13 http://dx.doi.org/10.1016/j.ejor.2009.06.004 .

Katsaliaki, K., & Brailsford, S. C. (2006). Using simulation to improve the blood supply chain. *Journal of the Operational Research Society, 58*(2), 219–227. doi:10.1057/palgrave.jors.2602195.

Korzilius, H., Raaijmakers, S., Rouwette, E., & Vennix, J. (2014). Thinking aloud while solving a stock-flow task: Surfacing the correlation heuristic and other reasoning patterns. *Systems Research and Behavioral Science, 31*(2), 268–279. doi:10.1002/sres.2196.

Kotiadis, K., Tako, A. A., & Vasilakis, C. (2014). A participative and facilitative conceptual modelling framework for discrete event simulation studies in healthcare. *Journal of the Operational Research Society, 65*(2), 197–213. doi:10.1057/jors.2012.176.

Lane, D. (1994). Modeling as learning: A consultancy methodology for enhancing learning in management teams. In J. Moorcroft (Ed.), *Modeling for learning organisations*. Portland, Oregon: Productivity Press.

Lane, D. (1995). On a resurgence of management simulations and games. *Journal of the Operational Research Society, 46*(5), 604–625. doi:10.2307/2584534.

Langley, P. A., & Morecroft, J. D. W. (2004). Performance and learning in a simulation of oil industry dynamics. *European Journal of Operational Research, 155*(16), 715–732.

Monks, T., Robinson, S., & Kotiadis, K. (2014). Learning from discrete-event simulation: Exploring the high involvement hypothesis. *European Journal of Operational Research, 235*(1), 195–205 http://dx.doi.org/10.1016/j.ejor.2013.10.003 .

Onggo, B. S. S., & Hill, J. (2014). Data identification and data collection methods in simulation: A case study at ORH Ltd. *Journal of Simulation, 8*(3), 195–205. doi:10.1057/jos.2013.28.

Petty, R. E., Briñol, P., & Tormala, Z. L. (2002). Thought confidence as a determinant of persuasion: The self-validation hypothesis. *Journal of Personality and Social Psychology, 82*(5), 722–741. doi:10.1037/0022-3514.82.5.722.

Pidd, M. (2002). Simulation software and model reuse: A polemic. In E. Yücesan, C. H. Chen, J. L. Snowdon, & J. M. Charnes (Eds.), *Proceedings of the 2002 winter simulation conference*. Institute of Electrical Engineers.

Robinson, S. (2001). Soft with a hard centre: Discrete-event simulation in facilitation. *Journal of the Operational Research Society, 52*(8), 905–915 http://www.jstor.org/stable/822951 .

Robinson, S. (2014). *Simulation: The practice of model development and use* (2nd ed.). London: Palgrave Macmillian.

Robinson, S., Nance, R. E., Paul, R. J., Pidd, M., & Taylor, S. J. E. (2004). Simulation model reuse: Definitions, benefits and obstacles. *Simulation Modelling Practice and Theory, 12*(7-8), 479–494.

Robinson, S., Radnor, Z. J., Burgess, N., & Worthington, C. (2012). SimLean: Utilising simulation in the implementation of lean in healthcare. *European Journal of Operational Research, 219*(1), 188–197. doi:10.1016/j.ejor.2011.12.029.

Robinson, S., Worthington, C., Burgess, N., & Radnor, Z. J. (2014). Facilitated modelling with discrete-event simulation: Reality or myth? *European Journal of Operational Research, 234*(1), 231–240 http://dx.doi.org/10.1016/j.ejor.2012.12.024 .

Rouwette, E. A. J. A., Bleijenbergh, I., & Vennix, J. (2014). Group model-building to support public policy: Addressing a conflicted situation in a problem neighbourhood. *Systems Research and Behavioral Science* in press. doi:10.1002/sres.2301.

Rouwette, E. A. J. A., Größler, A., & Vennix, J. A. M. (2004). Exploring influencing factors on rationality: A literature review of dynamic decision-making studies in system dynamics. *Systems Research and Behavioral Science, 21*(4), 351–370. doi:10.1002/sres.647.

Rouwette, E. A. J. A., Korzilius, H., Vennix, J., & Jacobs, E. (2011). Modeling as persuasion: The impact of group model building on attitudes and behaviour. *System Dynamics Review, 27*(1), 1–21. doi:10.1002/sdr.441.

Rouwette, E. A. J. A., Vennix, J. A. M., & Mullemkom, T. (2002). Group model building effectiveness: A review of assessment studies. *System Dynamics Review, 18*(1), 5. doi:10.1002/sdr.229.

Scott, R. J., Cavana, R. Y., & Cameron, D. (2015). Interpersonal success factors for strategy implementation: A case study using group model building. *Journal of the Operational Research Society, 66*(6), 1023–1034.

Shields, M. (2001). An experimental investigation comparing the effects of case study, management flight simulator and facilitation of these methods on mental model development in a group setting. In *Proceedings of the 19th annual international system dynamics conference.*

Shields, M. (2002). The role of group dynamics in mental model development: An experimental comparison of the effect of case study and management flight simulator under two levels of facilitation. In *Proceedings of the 20th annual international system dynamics conference.*

van der Zee, D.-J., Holkenborg, B., & Robinson, S. (2012). Conceptual modeling for simulation-based serious gaming. *Decision Support Systems, 54*(1), 33–45 http://dx.doi.org/10.1016/j.dss.2012.03.006 .

van der Zee, D.-J., & Slomp, J. (2009). Simulation as a tool for gaming and training in operations management – a case study. *Journal of Simulation, 3*, 17–28. doi:10.1057/jos.2008.8.