Experimental Evaluation of User Performance on Two-dimensional and Three-dimensional Perspective Displays in Discrete-Event Simulation

Ikpe Justice Akpan[a,*] Roger J. Brooks[b];
a. Department of Management & Information Systems, Kent State University, 330 University Drive, New Philadelphia, OH 44663, USA; b. Department of Management Science, Lancaster University Management School, Lancaster, LA1 4YX, UK.

Abstract

Several experiments were carried out to compare the impacts of using a two dimensional (2D) plan view or a three dimensional (3D) perspective view in discrete event simulation visual displays. The experiments measured the performance of participants in spotting errors, describing the model, and suggesting improvements to the system. The participants using the 3D perspective display performed much better in spotting errors, taking on average about one third of the time of participants observing the 2D display. They also did much better in describing the model. There was no significant difference in suggesting improvements although this may have been because this task was easy. Most participants preferred the 3D perspective view when asked to compare the displays. The experiments indicate that the detailed design of the visual display may have a considerable effect on some of the tasks in a simulation project and hence on whether the overall project is successful.

Keywords: discrete event simulation, visual display, model validation, model verification, two-dimensional display, three-dimensional perspective

1. Introduction

Three-dimensional (3D) visual display and Virtual Reality (VR) technology were first introduced in Discrete Event Simulation (DES) in the mid 1990's [7,20]. Since then, there has been tremendous increase in the recognition and implementation of 3D and VR applications, and its use for modeling activities within the DES community [15,30,39,41]. This recognition has led to strong consideration and adoption of 3D visualization and VR technologies as a new evolving modeling technique, and valid next steps in the advancement of DES [14,33].

Presently, most visual simulation applications and DES modelling activities involve the use of visual display in 2D and 3D. Unlike the 2D visual display, 3D visualization employs stereographic images to represent the model elements and other parts of the real system [32]. Numbers, charts and texts can also be utilized to display key statistics on the interface to complement the graphics, as also applicable in the 2D display.

_____

** Corresponding author. Tel.:+1 (330)-339-3391; Fax: +1 (330)-339-3321
E-mail address:* iakpan@kent.edu (I. J. Akpan)

In the DES literature, 3D visualization and VR is defined loosely to simply mean displays that give a three dimensional perspective view [30]. Although such displays can provide the ability to alter the viewpoint of the observer including being able to "fly through" the system, it is not really VR in the conventional sense as known in human-computer interaction (HCI) [23,32,33,36]. In HCI, VR represents variety of systems with different levels of immersion, including full-immersion into the virtual environment [32]. Immersive VR uses specialized hardware devices such as gloves and head mounted displays [27].

Generally, many of the benefits and claims of 3D visualization and VR in DES are similar to the advantages of visual interactive modeling and visual interactive simulation (VIS) [5,8-10], as decision support system [9,12,13,18] using 2D display [11]. The major difference is that, 3D display and VR provides greater enhancement to the benefits of visual simulation compared to the 2D display. For example, it is claimed that, 3D visualization and VR is more interactive, enhances generation of ideas about a simulated system [41], more effective for model testing, validation and verification [1,20,36], model credibility, and usability, and overall success of the simulation project [1,2]. While the importance of 2D visual display and 2D animation in DES is well established [8-11,17-19], the same cannot be said of 3D visualization and VR, with many ongoing debatable and unsubstantiated claimed benefits and costs.

We carried out a survey of simulation practitioners focusing on user perceptions of the differences between 2D and 3D displays [1], including perceived benefits and costs. But the survey results were not known at the time of the experiments reported in this paper. Generally the survey respondents with experience of both 2D and 3D considered 3D to be better than 2D for spotting errors, improving understanding of the real system, and communication with the client. One result that stood out was that 61% had the view that 3D resulted in a better solution for decision makers (with none saying that the solution would be worse) compared to using 2D. This effect is likely to be due to the 3D display facilitating better interaction with the client and helping the client to be more involved in the project. The respondents were also asked about the importance of five aspects of the visual display (3D perspective, realistic colors, detailed graphics, interactivity, accurate scale dimensions) in communicating with decision makers. A majority thought that each aspect was helpful with the responses being stronger from 3D users. However, there were a small number of comments in the survey that 3D could be a hindrance with unnecessary graphical detail making it harder to see important aspects of behavior.

Based on user opinions and evaluation results from 3D DES software vendors [1,7,8,31,39], it appears that the 3D visual display and the 3D animation can have a very significant effect on several tasks in a simulation project and on the overall outcome of the project, thus extending the benefits of the visual display. However, it is not clear precisely what factors of the visual display that 3D visualization can bring added benefits or the aspects of DES modeling process where 3D can affect the performance of developers or users of DES models. Therefore, the aim of the study was to investigate this through the use of experiments. Since 3D displays are increasing in popularity and prevalence in simulation, it was considered most useful to look at a 3D perspective

view. The experiments compared a 2D display with a 3D perspective display and were designed to look at the effect on the ability of users to spot errors in the model and to understand the behavior of the model (two of the benefits advocated for animation and visual display [5]), and make decisions on improving the business operations. We also examine the effects of the 3D perspective view on model credibility and acceptability. A display with just a 3D perspective view is sometimes described as 2.5D to distinguish it from displays with more capabilities, e.g. the ability to change viewpoint [39].

In the experiments reported in this paper, we are considering displays with just a 3D perspective view and we will use the term 3D display so as to give a clearer and simpler distinction in the text from the display with a 2D perspective. The experiments took place in 2004. Even though the use of 3D visualization in DES has become more widespread with many simulation packages developed considerably since then, the comparison of 2D and 3D perspective is still important and relevant in understanding the effects of this aspect of visual displays, especially as this aspect of empirical work yet remains overlooked.

The rest of the paper is organized as follows: Section two presents the theoretical background; Section three highlights the research hypotheses; section four explains the experimental design and method. The results of the experiments and test of hypotheses are presented in Section five. Section six discusses the implications of the study and limitations of the study, while Section seven summarizes the main findings and concludes the paper.

2. Theoretical Background

2.1 *Overview*

The dramatic advances in computer hardware and software, the revolution in the Internet and its multimedia front-end, and the World Wide Web have enabled the development and implementation of sophisticated user interfaces on diverse applications including DES as a decision support system [9,13,23,33,34]. While the benefits and costs of 3D visualization and VR in DES is still debatable, the application and implementation of these technologies are already well established in training simulations, entertainment industry, communication and media, and education with great successes [29,46]. For example, 3D games, 3D film productions and training simulation are very extensive [29,46]. In the area of training simulation, significant benefits have been recorded in flight simulation, military training and medical and surgical procedures [29,40,42].

The need for enhanced visualization in training simulation is often to make the display look realistic. On the other hand, such visual display may also be simplified to highlight the important features and it may show aspects that are not visible in the real system. Quarles et al. [29] describe a mixed reality approach to learning how to use an anesthesia machine. An existing virtual simulation model showed the workings inside the machine when the controls are altered, including showing the flow of invisible gases using colors. The mixed reality system was designed to make it easier for students to match the model with the real machine. It included altering the model display to match better with the real layout and using a magic lens to show the relevant components of

the model display on a hand held screen as the user looks at part of the real machine. The model was also linked to the real machine in that control movements on the machine were detected and applied.

The benefits of 3D display and VR in other fields as discussed above appears to have greatly influenced the quest to implement similar technologies in DES. It is claimed that, a realistic display may be necessary for simulation to have credibility as a cutting edge problem solving approach when compared against sophisticated computer graphics in other areas such as computer games [30], films [31], television, architecture, archaeology, etc. The public consumption of information such as news and current affairs is increasingly through video rather than the written word, and video is becoming more and more widespread on the internet.

2.2 *Visual Display in Discrete Event Simulation*

Most discrete event simulation (DES) applications and models have a visual display that provides a dynamic pictorial representation of the system being modeled. Our use of the term visual display follows the general usage in simulation [11,19] in referring to the use of images and graphics to represent the model elements in the simulation software. The visual display is usually dynamic in that, the model elements can move and change appearance at run time to indicate changes in states. Such a display may also include some text such as labels for the elements, as well as showing some key statistics through numbers or charts. This is generally termed animation. We distinguish the visual display from other aspects of simulation software and the simulation model such as computer code and detailed statistics. Simulation software also usually includes interaction in that the user can stop the model at run time to make changes to the model and then continue the run, thus making it possible for users to perform experimentation and analyses (e.g. "what if" analysis) [12].

As in most decision support applications [18,12,34], visual display has long been used in simulation through the implementation of visual interactive modeling (VIM) [3,9,17,19] and visual interactive simulation (VIS) [5,8,19,28]. The application of visual display has not only increased DES modeling and simulation activities, but revolutionized the modeling process [30], and remains an important part of a DES modeling till today.

2.2.1 *Benefits of Visual Display*
One benefit of visual display is the animation it provides, which can play a major role in achieving the overall objectives for most simulation projects. Simulation is generally used to model complex non-linear systems with many interacting components. Solutions are typically obtained by experimentation and it is often difficult to understand or anticipate precisely how the different parts of the system will affect the overall system performance. The ability to see the model running on the visual display can help considerably in understanding the behavior of the model and the system it represents. This can lead to new insights regarding ways of improving the system [39]. The visual display is important in other simulation tasks as well, such as model verification and validation [6]. Observing model behavior is a useful test as part of both verification and validation. Incorrect or unrealistic behavior may be noticed which would not have been spotted from a statistical

summary of the results. If an error is identified, then animation can enhance the debugging process to finding the root cause of the error [6,20]. This will often be done by running the model very slowly. The modeler can use the animation in each of these ways.

Another crucial role of the animation is that it enables other people to observe the model and the model runs, such as the client, decision makers, system operators and senior management. This allows them to make inputs into the modeling process particularly, as already described, during verification and validation, and experimentation, enabling their knowledge and expertise about the system to be used in these tasks. It can also help them to understand better both the system and simulation methodology. This is likely to improve communication between all parties involved in the project, facilitate acceptance of the use of simulation to tackle the problem, and increase model credibility [10,11,31]. The involvement of the client in the modeling process and their understanding of the process can be extremely important factors in the success of simulation or other Operations Research (OR) projects [9,35]. A good illustration of many of these benefits of animation in practice is provided by [20] in a case study of a construction simulation of operations to dig out and transport material. [5,20,41] describe several instances of how the simulation animation was used in verification, validation, and generating ideas for improving the system. [20] illustrate with the modeler seeing various incorrect truck movements which resulted in collisions or near misses, and the system expert noticing that the excavator dug out the material too quickly for the particular situation being modeled. Control rule ideas for the system were developed from the animation through observing the patterns and interactions of truck movements. [20] viewed that it would have been very difficult to find some of the errors and strategic ideas without the animation. This can also include examining statistics during the run and changing the view of the animation by zooming or altering the viewpoint [31]. In some cases interaction is built into the model coding so that the model and the animation stops at certain points and the user is asked to decide on the next actions to take to control the system. Such use of VIS is traceable to Hurrion's work in the late 1970s [10,11].

Some other benefits of visual display can be seen at the model building stage [1]. The graphical elements and visual display can often be used in constructing at least part of the model. This is typically achieved by operations such as selecting icons to add elements to the model and using the mouse, buttons or menus to connect the elements together.

Another area where visual display in DES can also provide great benefits is in the decision-making process and in facilitating the involvement of non-technical personnel in the simulation projects whatever their role. Engaging staff through model demonstrations can help with problem definition and setting objectives, model validation, and identifying ideas for improving the system that can be tried in experimentation. It helps the user, especially non-technical managers to experiments with the model to test different ways of organising, planning and running the system, and using the model results to predict the effect of making the changes in the real system which therefore helps to identify the best option [1,9].

Visual display also provides better understanding of the modelling process leading to a greater chance of the recommendations of the study being implemented [31]. This is applicable to wide variety of systems and operations including, manufacturing production lines, call centres, business processes, patient flows in hospitals, and road traffic junctions simulation. It can also increase communication and coordination within the organisation by enabling staff to see a broader view of the problem beyond their specific area of responsibility [40]. Therefore, having a model display that can be understood easily by the client can be a very important part of a simulation project being successful, even though it doesn't affect the logic and behaviour of the model.

It is easy to overlook the role of the animation in simulation as it does not affect the behavior of the model, and it often receives little attention in simulation textbooks. Consequently it can seem relatively unimportant compared to the other simulation tasks such as conceptual modeling, data collection, model coding, verification and validation, and experimentation. However it can make the difference between a simulation model being a black box that, at least to the client, is slightly mysterious, and being a transparent and understandable technique. This is vital though, as visual animation is one aspect in which simulation differs from many other operations research techniques. According to Rohrer [31], the visual display is particularly important because human beings are naturally good at processing visual information. However, [31] also set out reasons why visual display does not receive enough attention in DES, namely the reading and writing emphasis in the education system, graphics being implicitly regarded as a game rather than serious analysis, and graphics being considered as requiring a lot of effort for at most small gains [1].

The benefits of the animation are likely to vary depending on the specific model application. Some systems may be difficult to represent because they are too large or have too many elements or because their elements do not have a simple and natural way of being represented. On the other hand, in some applications, such as facility layout problems [e.g., 39] or the movements of people in crowds [e.g., 22], being able to visualize the system can have a special importance.

2.2.2   *Limitations of Visual Display*
Of course, the visual display has its limitations. In a large system then it may be difficult to observe the whole system on the screen and to identify the important factors. The use of numbers, statistics and charts is usually essential in the proper analysis of experiments and the animation is often turned off during experiments so that the simulation runs faster. Since the animation can be powerful there is also the possibility for it to be misleading and to give a false confidence in the model simply due to impressive and high quality graphics.

2.3 *Benefits of 3D Visualization and VR Over 2D Visual Display*

The benefits of visual display as discussed in Sections 2.2 apply to both 2D visual display and 3D visualization in DES. However, this section identify the aspects of discrete event simulation and modeling where 3D can make

better impacts compared to the 2D based on several claims in the literature and user opinions [1,2]. Thus, it is claimed that 3D enhances of improves the following benefits of visual display:

- Generation of ideas about the system being modeled and helping users to "gaining insights and relaying knowledge" [36,41]. Also, 3D graphics and 3D animation uses true to scale elements making simulation models easy to understand.

- Communication with clients and help in presenting the model and solutions [7,39,41].

- Model testing, validation and verification [1,40].

- Better problem definition that is easily agreeable by all stakeholders of the simulation project than 2D display [1].

- Model credibility, acceptance and usability. The 3D models easily convey results and make any recommendations arising from the simulation more convincing and credible, and also lead to increased confidence in the model [24, 40].

- Improving the Quality of Managerial Decisions. 3D models are easily comprehensible, providing clear information to decision-makers or model users, and improving understanding of the modeled system reasonably [1,39,45].

- Bridging the communication gap between model developer and management or non-technical personnel, 3D visualization can become a catalyst for resolving complexities in the simulation models, and improves the quality of decision-making [39].

3. Research Hypotheses

The opinions of DES practitioners and users of simulation solutions and DES models [1], the evaluation results of 3D software by vendors [7,39], and the outcomes of systems modeling and simulation using 2D display and 3D visualization [20,30,36,41,45] all indicate that, 3D display could bring significant benefits on DES modeling activities. This experimental study aimed at comparing users' performance on discrete event simulation activities using the 2D display and 3D perspective view, and identifies aspects of simulation and modeling activities where 3D display can bring added benefits. Specific modeling activities addressed in this study include, model validation and verification, generation of ideas and insight into the modeled system/operation, model acceptability and credibility and decision-making. The reason for focusing on these modeling activities for the experiments was based on two reasons; first, the claimed benefits of 3D display in the listed modeling activities appear very strong; second, the participants in the experiments simply watched the DES models at runtime and observed the animation, with no interaction with the display due to limited knowledge of subjects in discrete event simulation, especially in 3D modeling. However, a similar experimental study that examined the efficacy of VIS compared with non-visual interface [11] used subjects with similar characteristics as ours.

The research hypotheses formulated in this study as presented in Section 3.1 – 3.5 were based on the various claims and benefits as listed above.

3.1 *Model Validation and Verification*

Model Validation is a process of determining whether a simulation model is an accurately represent the actual system based on the specific objectives of the simulation project [20]. The purpose is to ensure that the simulation model accurately mimics the real-world system from the perspective of its intended uses, and includes activities such as checking for errors in the model and correcting them. The model is tested until we have sufficient confidence in it to use for decision-making. [20,21] emphasize validating models using the visual display. [1,36] present strong indications about the efficacy of 3D display for model validation compared to 2D.

We designed the following hypothesis to test the claims that 3D visualization is more effective than 2D display in validating DES model. We defined effectiveness in this context to mean that, 3D display makes it easier to spot errors in DES model, measured by the number of subjects who are able to detect specific errors, and the time taken to spot such errors.

- *Hypothesis 1:*

We therefore define the null hypothesis $H1_0$ and the alternative hypothesis $H1_1$ as follows:

    $H1_0$: 3D visualization does not make it easier to spot errors in DES model than 2D display.

    $H1_1$: 3D visualization makes it easier to spot errors in DES model than 2D display.

3.2 *Model Understanding and Generation of Ideas*

We assessed the effect of 3D visualization and 2D display on the ability of observers of the model to understand the models and generating ideas to make decisions.

We designed two hypotheses (hypotheses 2 and 3) to test the claims that 3D visualization can enhance users' understanding of the model and the simulated operation, and generation of ideas towards improving the quality of managerial decisions.

- *Hypothesis 2:*

The null hypothesis $H2_0$ and alternative hypothesis $H2_1$ were defined as follows:

H2₀: 3D visualization does not result in a better understanding of the model than 2D display.

H2₁: 3D visualization results in a better understanding of the model than 2D display.

- *Hypothesis 3:*

On the claims that 3D display can enhance generation of ideas towards improving the quality of managerial decisions [1], we defined the following null hypothesis $H3_0$ and the alternative hypothesis $H3_1$:

H3₀: 3D display does not enhance generation of ideas for improving business decisions than 2D.

H3₁: 3D display enhances generation of ideas for improving business decisions than 2D.

3.3 *Model Acceptability*

Two hypotheses (hypotheses 4 and 5) were designed to test the claims that 3D visualization enhances model credibility and acceptance. Credibility implies adjudging the model as correct representation of the real system based on the interaction between model elements, and the visual representation of the actual operation.

- *Hypothesis 4:*

The fifth hypothesis tested the claim that using 3D display enhances model acceptability compared to using the 2D display. The null hypothesis H_0 and alternative hypothesis H_1 were defined as follows:

H4₀: 3D visualization does not improve model acceptability than 2D display.

H4₁: 3D visualization improves model acceptability than 2D display.

4. Research Method

This research adopted experimental strategy to test the hypotheses developed in Section 3 above, as discussed in this section.

4.1 *Experimental Design*

We adopted simple between-subjects design [25] for all the experiments discussed in this paper. The experiments followed a randomized block design with one factor (display type) and two levels; 2D display and 3D perspective view (figure 1). The dependent variable was the performance of users, e.g. spotting of errors (experiment 1), model understanding (experiment 2) and model credibility and acceptability (experiment 3). Each subject performed the tasks using either the 2D display or 3D display. This helps to overcome any experimental errors that can arise from "ordering effects" or "learning effects" [25] should participants perform the tasks on both displays. Further, to mitigate the 'learning effect' on the experimental results (performance of tasks between subjects), the tasks were assigned to all participants in the same order. For example, all participants were made to spot the routing error first, followed by the wrong part error. This helps to overcome any possible confounding errors such as differences in performance due to alternating the order of activities

undertaken between participants. The block sampling design also helped to eliminate any possible biases caused by different backgrounds of subjects. For example, of the 26 participants with background in simulation and other operations research disciplines, we assigned 13 to perform the task using 2D display and the other 13 to use the 3D perspective view (table 1). Altogether, all participants were divided into two groups, 31 subjects in each group doing the experiments on 2D and 3D displays respectively.

Other actions that were taken to control the experiment, processes and manipulate variables of interest were as follows:

- The maximum time allowed for each task was set prior to the commencement and was maintained throughout the experiment.

- The resources used for the experiment including equipment such as the lab used were setup and planned prior to the experiment. The reason was to control any possible external factors (such as noise level) that can affect the performance of participants and the eventual outcome from the experiment.

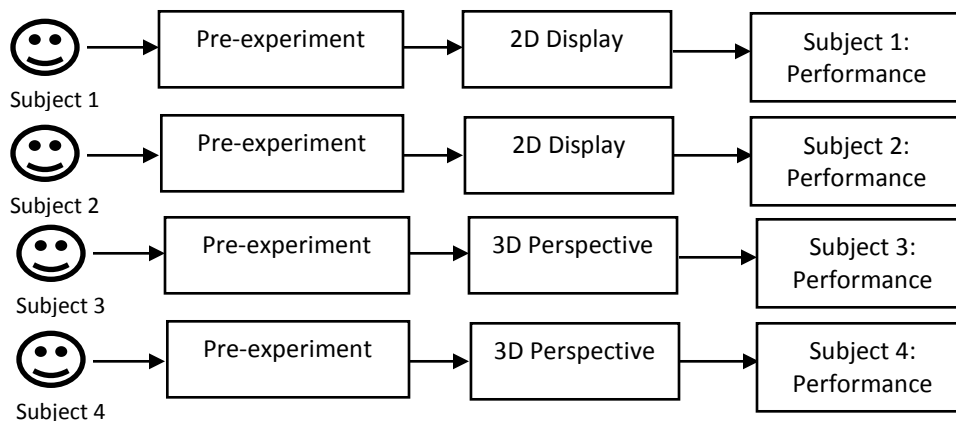- All the participants performed identical tasks.

Fig. 1. Between Subject Experimental Design

4.1.1 Pre and Post Experiment Activities

Several activities were carried out before the start and at the end of the experiment (figure 1) as discussed below.

- **Experimental Questionnaires**

Participants completed three short questionnaires/forms during the experiment namely, the background and psychological questionnaire, forms to document the results of the first experiment and question sheets for the second experiment (the main experiments). No questionnaire was completed for the third experiment as the experimenter documented the relevant data [2].

Prior to the start of the first experiment, each subject was handed a written instruction about the various activities to be completed during the 30 minutes session that each participant spent for the three experiments.

Each subject had an opportunity to ask questions (if any) after reading the instructions/guidelines. They were also given the short psychology questionnaire and documentation forms to complete.

- **Psychology questionnaire**

The psychology questionnaire was designed to assess the participant's cognitive learning styles [25,43,44] through the use of visual language and visual numeric symbols. Subjects who learn using visual language can remember and understand things better by seeing rather than by hearing. When words are read to them, they are more likely to write down those words on paper or boards. Similarly, those who learn through visual numerical symbols are more likely to remember and understand better if numbers are written on paper or board, or in a book. Also, they do not seem to need much oral explanation [43,44].

For visual language the participants were asked to rate themselves on a scale from 1 (least like me) to 5 (most like me) for the following statements:

- I remember what I see better than what I hear
- I understand information by visualizing pictures
- I use different colors to highlight, select and organize when writing or studying.
- I make study notes using visual drawings, spacing, symbols, etc.
- I recall written information by visualizing text pages, notes or study cards
- I would rather read a story than listen to it read.

For visual numeric symbols, the same rating scale was used with the following statements:

- I understand math/statistics problems that are written better than one that I hear.
- A graph or chart of numbers is easier for me to understand than numbers written.
- Written math problems are easier for me to solve that oral ones.
- I learn better by reading than listening.
- Seeing a number makes more sense to me than hearing a number.
- I use study cards with written information and review them by writing to reproduce the information.
- I make recall cues as visual as possible.

The background questionnaire was completed before the beginning of the experiments, and included information on gender, academic background, level of study, the use of visual language and visual numeric symbols. Each subject completed the documentation forms at the end of the first experiment, and the questions and answers sheet for the second experiment. The documentation questionnaire for the first experiment contained information on academic background, level of study and any experience of 2D and 3D simulation. Subjects were also asked

to rate their understanding of the simulation models used for the first and second experiments respectively, the usefulness of the graphics in understanding the simulation model, and indicate how difficult or easy the errors were to spot. They were also given opportunity to make any general comments about the experiments. The form used for the second experiment asked the subjects to answer some questions about their understanding of the model and recommendations for decision-making about the simulated system.

4.2 *Participants*

The subjects were students and staff / faculty from Lancaster University (U.K.). The sessions took place over a two-week period and participants signed up for a particular time of their choice. The participants were paid for taking part except for two staff / faculty who volunteered to do the experiment free of charge. It was not known at the start exactly how many participants there would be as participants could continue to sign up until the end of the two week period. In the end there were 62 participants.

No restriction was placed on the backgrounds of participants although the experiment was mainly advertised within the Management School and in the student union building. A broad range of participants from different fields of study was considered advantageous because people involved in simulation projects can have very varied backgrounds. Although the modeler will often (although not always) be a simulation specialist, other people such as management and system operators will typically be involved, and they may be unfamiliar with simulation and may not even have much knowledge of quantitative methods.

In the detecting errors experiment, participants were mainly allocated in alternate blocks of one, two or three participants (a morning or afternoon schedule of six participants might be allocated three 2D then three 3D participants). This was to ensure equal numbers in each group given that the total number of participants was not known until near the end of the experiments. The participants chose their own time in the schedule and this provides a pseudo random element to the allocation process. Therefore, a manual allocation rather than one using, say, random numbers was considered acceptable.

In addition, factors that might affect the performance of the participants were considered and it was anticipated that an important factor might be the quantitative and analytical abilities of the participant. The best indication of this in the information available was the subject of study of the participant. To try and reduce the effect of this factor, the cumulative number of participants from each general subject area was recorded as the experiments progressed and participants from each subject area allocated equally to the two groups as far as possible. The same principles were applicable for the model understanding experiment.

Table 1 shows the characteristics of the participants. The groups have similar numbers in the different subject areas due to the chosen allocation method, as described above. There is some subjectivity in deciding which subject area the participants belong to and, in fact, this was changed more than once during the analysis and checking of the data. For example, some students were studying the management of information technology and

so could be included either the business or computing categories (in the end, business was chosen as the most suitable, particularly as it is a Management School course). The most important areas of study to allocate equal number of subjects as much as possible are the simulation PhD students, who should have particularly good skills for the tasks, and the arts, humanities and social science students who probably have the least training in analytics. The students in the other categories should have good general quantitative skills, with most having studied several courses in statistics and modeling.

In the other characteristics, both groups have very similar numbers with some experience of simulation, which provides further reassurance of similar skill levels between the groups. The 3D groups have a higher proportion of males. The 3D groups have a slightly higher education level overall than the 2D groups, with more postgraduate and staff participants and fewer undergraduates (although for the model understanding experiment the 3D group has fewer PhD students). A higher education level and, presumably, greater experience, would be expected to be of some advantage. However, overall we have good confidence that the characteristics of the two groups are similar and we are not aware of any differences that might have a major effect on average performance.

Table 1.
Characteristics of the participants

Characteristic	Detecting errors		Model understanding	
	2D	3D	2D	3D
Total number	31	31	31	31
Gender				
Male	16	22	17	21
Female	15	9	14	10
Study level				
Undergraduate	11	4	9	6
Masters / postgraduate	5	10	5	10
PhD	15	15	17	13
Staff	0	2	0	2
Current subject area of study				
Simulation PhD	2	3	3	2
Other Operations Research	11	10	11	10
Other Business and Management	11	12	11	12
Computing, Information Technology	2	1	1	2
Engineering, Science	2	1	2	1
Arts, Humanities, Social Science	3	4	3	4
Experience of simulation				
Experience of 2D simulation	19	21	20	20
Experience of 3D simulation	1	0	1	0

4.3 *Experimental Procedures and Processes*

The applicable procedures and processes to all the experiments carried out in this study are discussed in this section. Any processes or procedures that relate specifically to any of the three experiments are presented in the relevant sections. The experiments carried out and their sequence was as follows:

i. Detecting errors experiment (car assembly model).

ii.   Model understanding experiment (bank model).

iii. Subjective Model assessment (car assembly model).

Each session of about 30 minutes involved only one participant at a time. Prior to arrival, each participant was randomly assigned to perform the tasks in either 2D or 3D simulation models using precise sampling technique [25].

Prior to the start of the first experiment, each subject was handed a written instruction about the various activities to be completed during the 30 minutes (max) session that each participant spent for the three experiments. Each participant could ask questions (if any) after reading the instructions/guidelines about the experiments prior to commencement, and also completed pre-experiment questionnaires as discussed in Section 4.1. The complete processes and procedures for all the experiments are shown in figure 2.
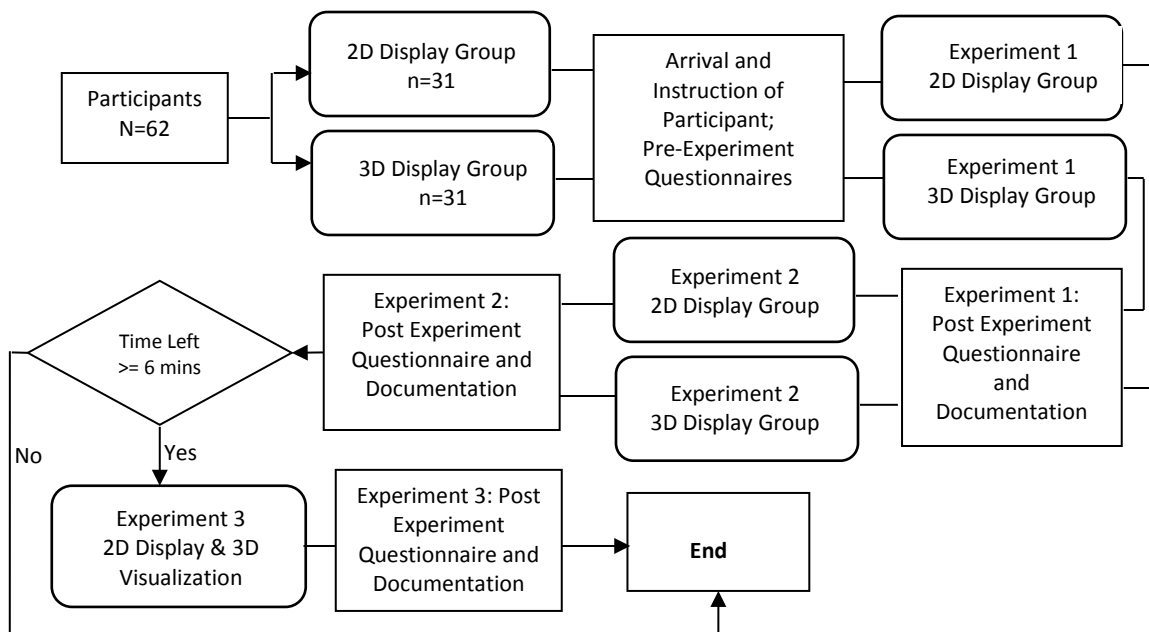
Fig. 2. Procedures and Processes for the Experiments

4.4 Detecting Errors Experiment

4.4.1   Car Assembly Model Description. The model used for this experiment was a simulation of a fictional car assembly production line. The model was chosen and designed to meet the following criteria:

- The model is based on a product and a situation that most people will understand and be reasonably familiar with so as to reduce the differences between subjects from different backgrounds.

- The model is fairly simple and so is understandable in the short time available for the experiment.

- The model has sufficient dynamic behaviour in the display so that the task set is not trivial.

The model represents a simplified car assembly plant which works as follows. There is a constant supply of three components (bodies, doors and tires) into the factory. The first process is that each component is tested (on separate machines) with rejected parts being scrapped. Next is the assembly process which assembles 1 body, 2 doors and 4 tires into a complete car. Then complete car is tested. Accepted cars are shipped out and leave the model. Rejected cars are sent to be reworked. Most cars are reworked successfully and are sent to be re-tested. If the rework is unsuccessful the car is scrapped. The model contains six machines and the parameters for the machines are shown on Table 2. Also, figure 3 shows the complete model logic of the car assembly operation.

Table 2.

Parameters for the machines in the car assembly model (m = mean, s = standard deviation, all units are in minutes)

Machine	Cycle time	Rejected	Breakdown interval	Repair time	Labour (for machine repairs
Test tyre	12.5	10%	-	-	-
Test body	40.0	5%	Exponential, m = 600	Lognormal m = 2, s = 10	Inspection Technician
Test door	25.0	5%	Exponential, m = 60	Lognormal m = 2, s = 10	Inspection Technician
Assemble car	Uniform 35.0 – 45.0	0%	Exponential, m = 57	Lognormal m = 2, s = 9	Assembly Technician
Test car	25.0	40%	Exponential, m = 180	Lognormal m = 2, s = 10	Inspection Technician
Rework car	35.0	2%	-	-	-

Most machines break down. The breakdowns were set to occur often so that they would occur several times in the short period of the experiment. The repair times are short and so there is not much effect on the system. The purpose of including breakdowns was to add a bit more complexity to the animation and the main effect in the animation was that the labour icons briefly moved to the relevant machine while the repair took place.

The components move between the machines on conveyors. All conveyors are queuing conveyors, which means that if the part at the end of the conveyor has to queue, the parts behind it continue to move. Each conveyor has an index time of 1.5 minutes (this is the time for the part to move forward one space).

The model was built in WITNESS 2003 software. Two versions of the model were built which differed only in the visual display. The two versions therefore behave in an identical way. One visual display is a 2D model showing an overhead plan view of the car assembly plant (Figure 10 - the correct version of the model used for experiment 3). The other is a display which gives a 3D perspective view of the plant (Figure 11). The displays were designed to be as similar as possible apart from the perspective and so similar icons and colours were used in both displays. During the experiment the run speed of the models was set so that they both ran at the same speed. The run speed was about 1200 seconds in real time, and so 1 second of real time corresponded to about 20 minutes of simulation time.
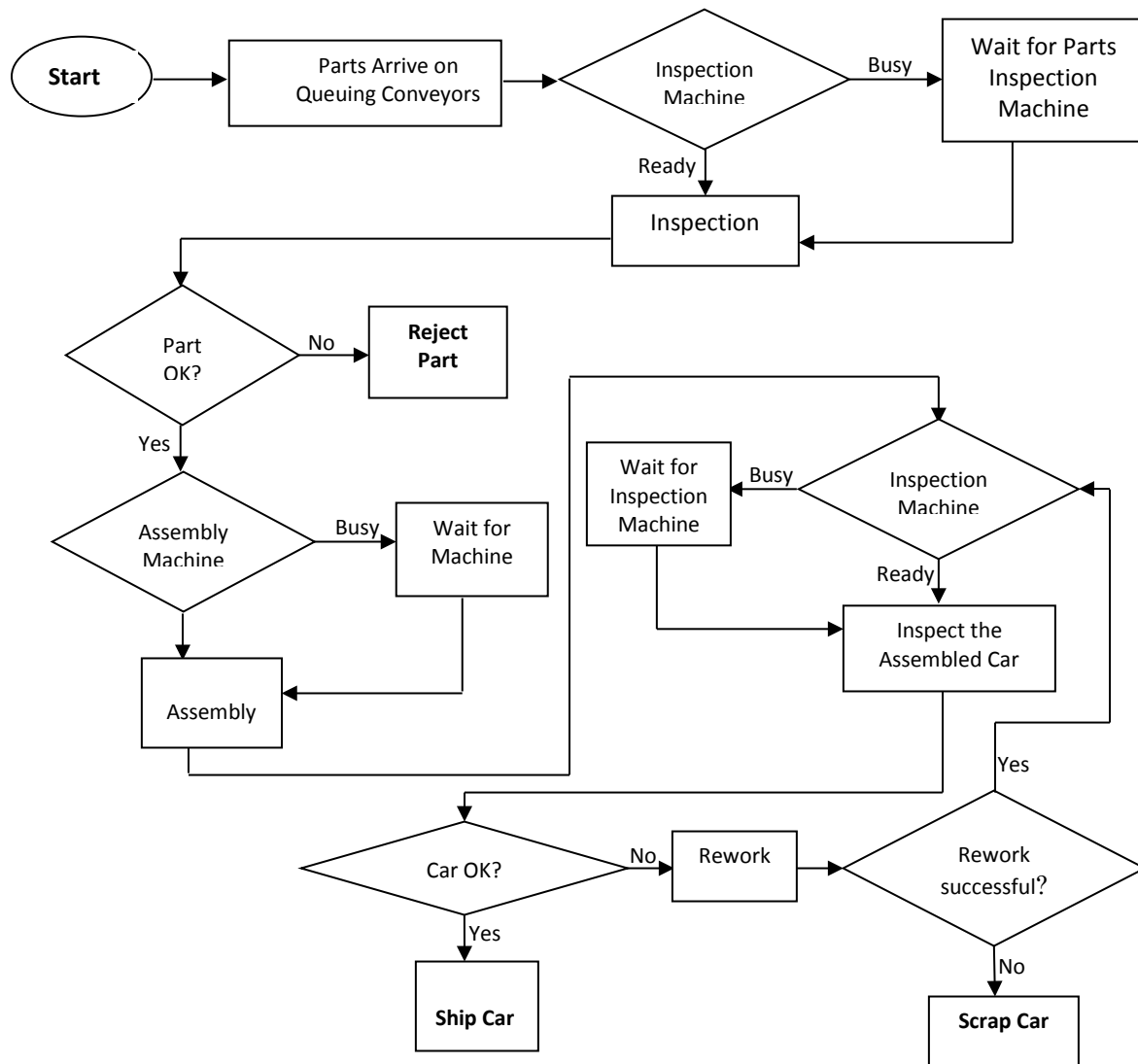
Fig. 3. Model Logic of Car Assembly Operation

4.4.3   *Errors Used in the Experiment.*

In choosing the errors to include in the model, the criteria were that the error should be one that could occur in practice, that can be spotted from the visual display, and that is not too easy or too difficult to spot. A trial experiment was conducted using 10 participants to get feedback both on the experimental procedure and on possible model errors. Four model errors were included in the trial experiment (including one whose format was revised). Based on the performance of the subjects and their feedback, two of these errors were chosen for use in the final experiments [16].

The two errors used were a routing error and an assembly error. The errors were included separately with the experiment run firstly with the model containing just the routing error, and then secondly with the model containing just the assembly error.

The routing error was an incorrect routing of the reworked assembled cars. In the correct version of the model assembled cars that fail the inspection (40% on average) are sent to the rework machine. After the rework operation, most reworked cars (98% on average) are sent back to the car inspection operation, with only a few (2% on average) being scrapped and sent down the conveyor to the scrapped location. In the incorrect version of the model the cars that fail the inspection are scrapped without being sent to the rework machine. This means that they go along the conveyor past the rework machine (Figures 3 and 4) to the scrapped destination. There are several possible ways to spot this error from the visual display as follows:

- The cars sent to the rework machine go straight past the machine to the scrapped destination.

- The rework machine is idle all the time.

- The conveyor on which the reworked cars are sent back to the inspection operation is always empty.

- The number of scrapped cars compared to the number of shipped cars shown on the model display is a ratio of about 40:60 rather than the expected ratio of about 2:98.
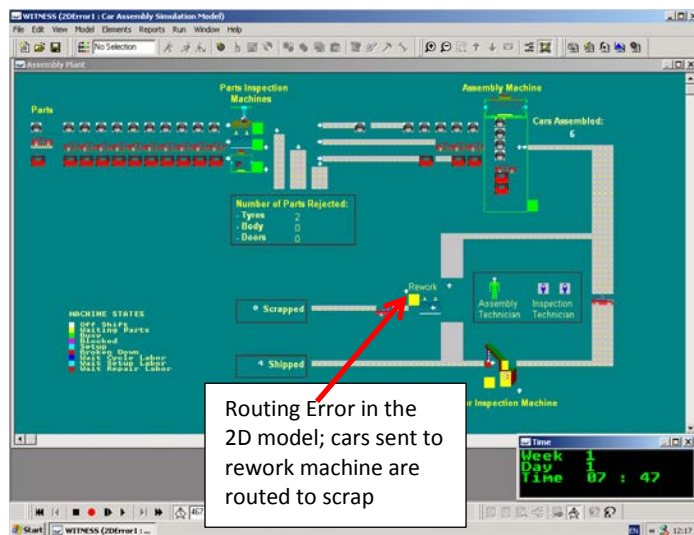

Fig. 3. Screen Shot of 2D Display of the Car Assembly Model at a Runtime, showing the Routing Error (the rework machine was idle, all cars sent to it were immediately scrapped).
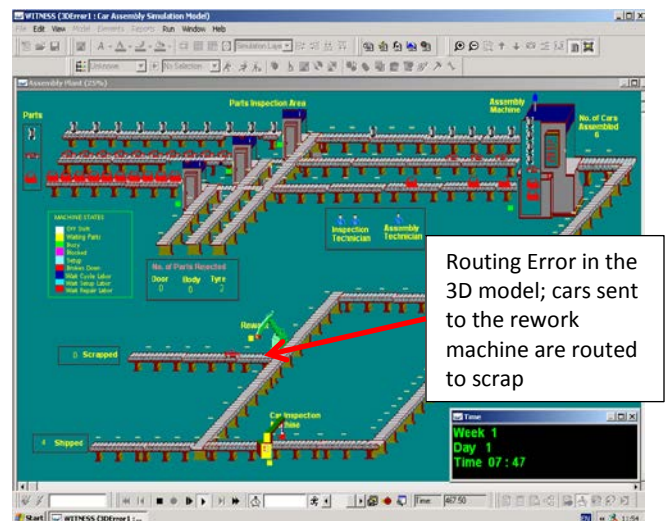
Fig. 4. Screen Shot of 3D Display of the Car Assembly Model at a Runtime, showing the Routing Error (the rework machine was idle, all cars sent to it were immediately scrapped).

In the assembly error, the incorrect number of components was used for the assembly operation. The assembly machine should have as its input four tires, two doors and one body. The error was that the machine takes five tires instead of four. Ways in which this error can be spotted are:

- The display shows the parts at the assembly machine during the assembly operation and there are five tires instead of four (this error is shown in Figures 5 and 6).

- The system is unbalanced with the assembly machine often idle and queues building up of components.

The task to spot the routing error was carried out first, followed by the assembly error. The same order was maintained for each subject. This is because there will be some learning effect during the first experiment as the subjects become more familiar with the model. Changing the order of the procedure between participants would have added an additional unwanted factor. When trying to spot the second (assembly) error, the participants were more familiar with the model compared to the routing error.
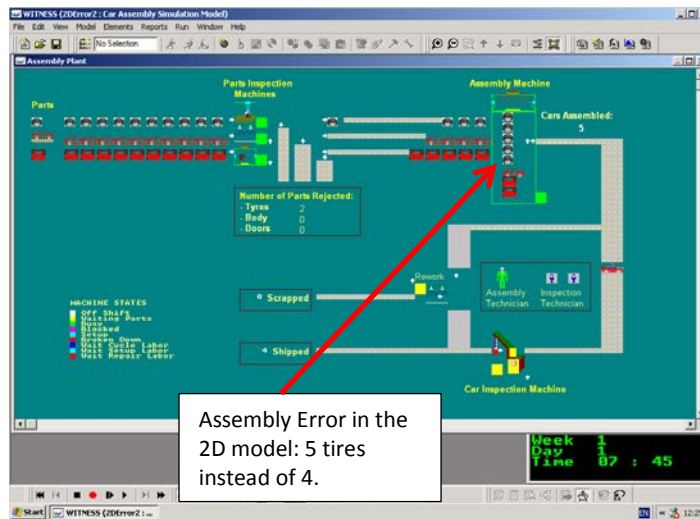

Fig. 5. Screen Shot of 2D Display of the Car Assembly Model at a Runtime of 465 Minutes in Simulation Time (about 23 Seconds in Real-time), showing the Assembly Error (5 tires at assembly machine).
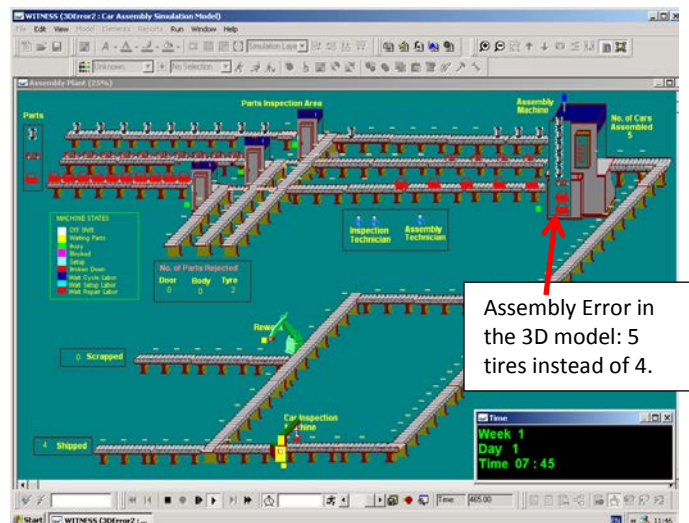
Fig. 6. Screen Shot of 3D Display of the Car Assembly Model at a Runtime of 465 Minutes in Simulation Time (about 23 Seconds in Real-time), showing the Assembly Error (5 tires at assembly machine).

4.4.4 *Procedure for the Detecting Errors Experiment.*

The procedure used for the detecting errors experiment was as follows:

a) Read model description. The subject was given a two page handout consisting of a list of the tasks that will take place during the experiment, a brief written description of the model and a screen shot of the version of the model (2D or 3D) that they will be using. The subject was given 5 minutes to read the handout and was allowed to ask the experimenter any questions about the information in the handout.

b) Routing error. The subject was told that the model contained an error. The subject was asked to watch the model as it runs and to try and spot the error in the model. They were told that when they thought they had seen the error that they were to state what it was. The experimenter would then tell them whether it was correct. If not correct then the model would continue to run and the subject should continue to try and spot the error. The experimenter then opened and ran the model containing the routing error. The experiment proceeded as just described. The experimenter recorded the time when the subject found the error. The maximum time limit for spotting the error was 10 minutes.

c) Assembly error: It followed the same process as step 2 above, using the model containing the assembly error.

d) Questionnaire. The participant completed a short questionnaire containing questions to record their background information (gender, level of study and their previous experience of simulation), two questions about the ease of understanding of the model, and some questions about spotting the error. The time the error was spotted was also recorded on this form.

4.5 Model Understanding Experiment

4.5.1   Bank Model Description.

The model used for this experiment was based on bank customer service operation. As with the car assembly model, this model scenario was chosen since it is familiar to most people, is a simple model, and the display has several dynamic elements. The bank in the model has two types of staff for different services. Tellers provide standard services such as deposits and withdrawals. Enquiry staff were responsible for any other issues such as setting up new accounts.

There are four tellers and three enquiry staff with a single teller queue and a single enquiry queue. Customers arrive at random with a negative exponential inter-arrival distribution with a mean of 1.5 minutes. They join either the teller queue or the enquiry queue, with on average 90% joining the teller queue. The maximum queue size is 10. The teller service time is a triangular distribution with minimum 5, mode 15, and maximum 25. The enquiry service time is also triangular with minimum 5, mode 10, maximum 30. After being served, some customers require the other services and so join the other queue. This applies to about 6% of teller customers and 50% of enquiry customers. The system was deliberately designed to show a problem with the queue for tellers quickly building up to maximum. The different customer routes also provide some complexity to the animation. Figure 7 shows the model logic of the bank customer services operation. As with the car assembly model, the bank model was built with 2D and 3D displays as shown in Figures 8 and 9. Again, similar icons and colors were used to make both displays as similar as possible apart from the perspective view.
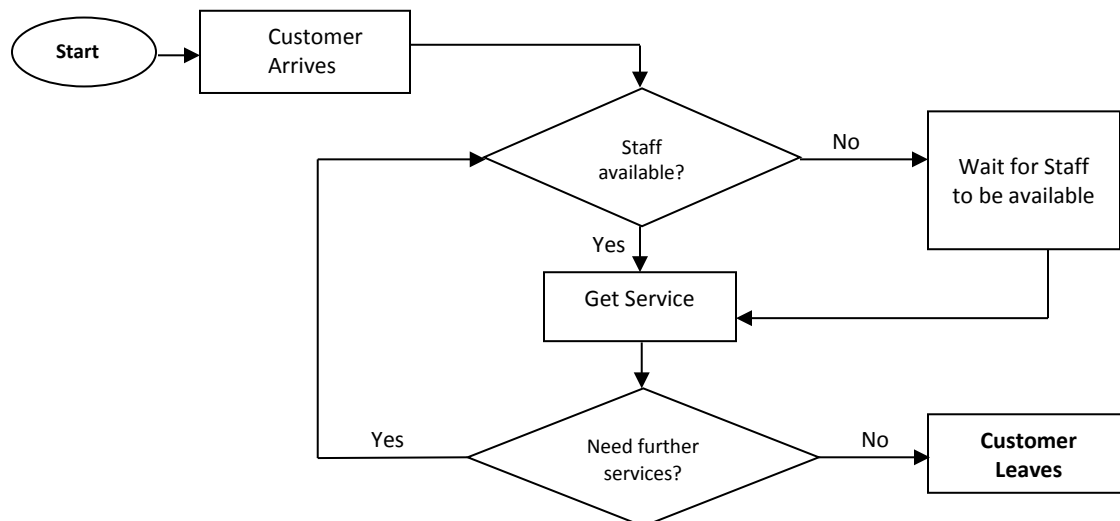
Fig. 7. The logic flow of the bank customer services model

4.5.2   Procedure for Model Understanding Experiment.

The procedure used for the model understanding experiment was as follows:

a) The experimenter explained the procedure of the experiment to the participants and gave them a handout explaining the experiment and the two questions they would have to answer. The participants read the questions but were not provided with information about the model.

b) The participants observed the model for 2 minutes of real time (101 minutes of simulation time).

c) The participants wrote the answers to two questions: (i) Please describe as fully as possible how the system works; (ii) The bank wishes to provide good service without incurring excessive cost, what changes to the system would you recommend and why?

d) Questionnaire. The participants provided ratings for two questions about the ease of understanding of the model (the same questions as for the detecting errors experiment).
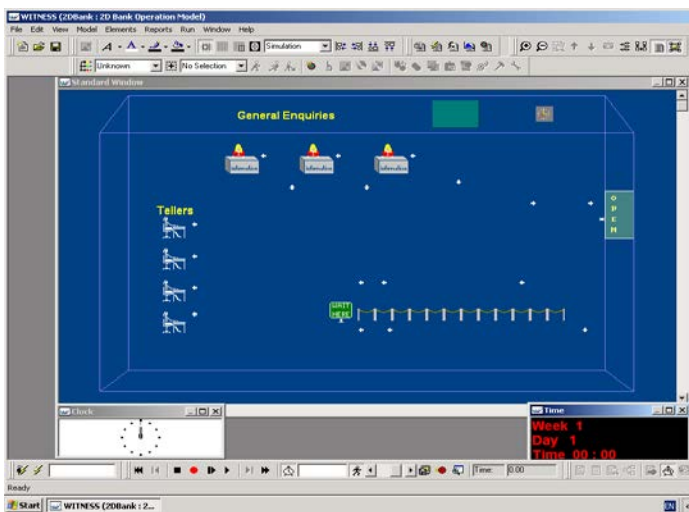

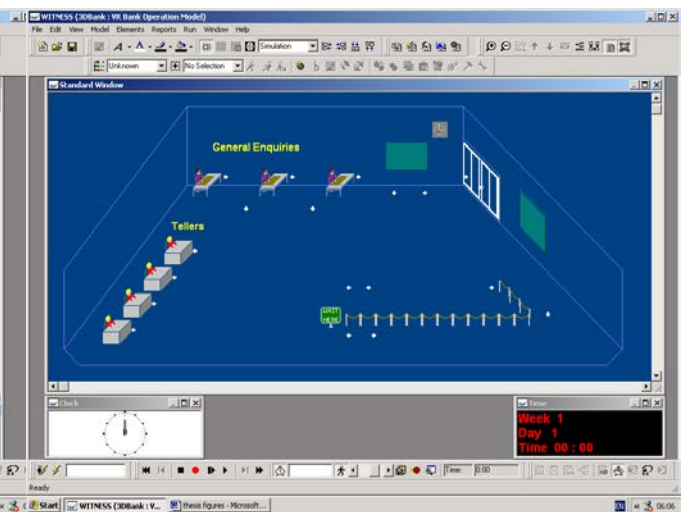Fig. 8. Screen Shot of 2D Display of the Bank Model.   Fig. 9. Screen Shot of 3D Display of the Bank Model

4.6 *Subjective Model Assessment*

If there was time left (at least 6 minutes) after completing the first and second experiments, the subject was assigned to undertake the third experiment using the car assembly model. This time the model did not contain any errors (figures 10 and 11). This was simply a subjective comparison of the two displays which consisted of the participants observing both the 2D and 3D displays and then answering two questions as to which one they preferred and why. The detailed procedure was:

a) A brief verbal reminder was given of the model description (the participants had already seen the model in the detecting errors experiment), as well as explaining the procedures for this experiment.

b) The participant observed the model as it was run for two minutes of real time with the display that was not used by the participant in the detecting errors experiment.

c) The participant observed the model as it was run for two minutes of real time with the display that was used by the participant in the detecting errors experiment.

d) The participant selected which display they preferred for two questions asking about different aspects of the model.
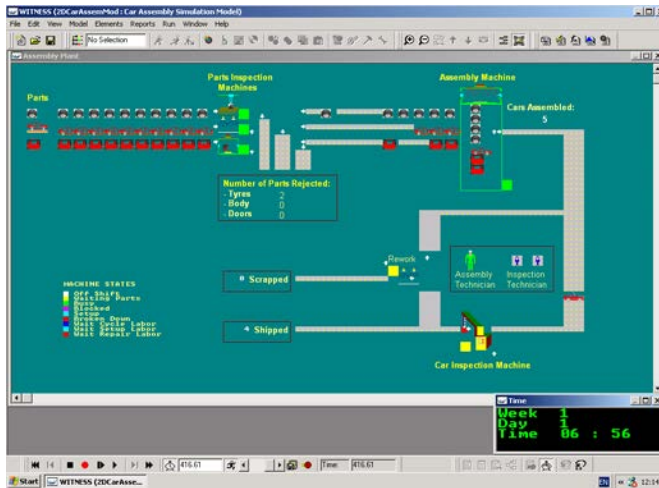

Fig. 10. Screen Shot of 2D Display of the Correct Version of the Car Assembly Model at a Runtime of 416.61 Minutes in Simulation time (about 21 Seconds in Real-time)
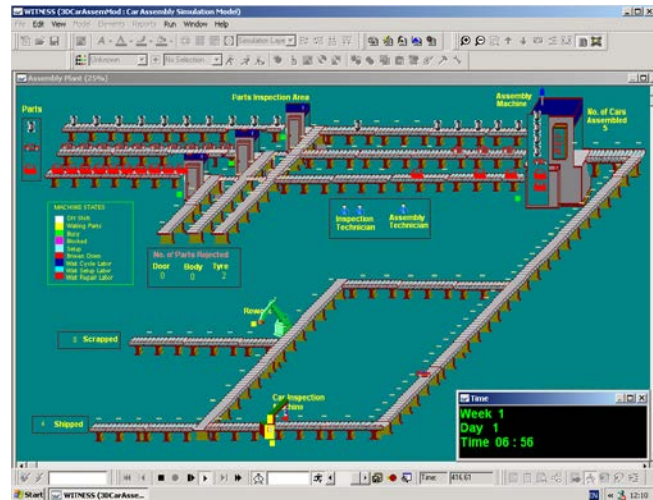
Figure 11. Screen Shot of 3D Display of the Correct Version of the Car Assembly Model at a Runtime of 416.61 Minutes in Simulation time (about 21 Seconds in Real-time)

5. Results and Hypotheses Testing

The data generated from the pre and post experiment questionnaires and the three experiments were analyzed using descriptive statistics, Chi-Square, and binomial tests. These results are presented in this section.

5.1 The Characteristics of Participants

The background and characteristics of participants include cognitive learning styles, level of education, area of study and experience in simulation. This was considered necessary for diagnostics in view of its possible effects on subjects' performance.

- **Cognitive Learning Styles**

The cognitive learning style of participants was based on visual language and visual numeric symbols as discussed in Section 4.1.1. The rating score in each category was added up and multiplied by a scalar to give the total percentage score, which determines the learning style of each participant. For example, participants with a visual score of 80% and over are considered as major visual language and visual numeric symbols learners,

scores of 60% to 79% as minor and less than 60% are classified as negligible [44, 47]. For those who carried the error detecting experiments using the 2D display, 10, 18, 3 were major, minor and negligible visual language learners, and 18, 12, 1 of major, minor and negligible visual numeric symbols learners; the 3D group were: 12, 18, 1 and 19, 12, 0 of major, minor and negligible visual language learners and visual numerical.

Figure 12 shows how the participants classified as major, minor and negligible visual language learners spotted the routing error in 2D display and 3D perspective view. As expected, most of the major visual language learners (70%, 83%) spotted the routing error by the help of 2D and 3D animation respectively. On the other hand, a good number of those classified as minor visual language learners (39%) relied on statistics to spot the error in the 2D display. The other 39% in the same category spotted the error using 2D animation, while 17% spotted the error with the help of statistics and 2D animation, and about 5% of this category could not spot the error in the 2D display. All categories of visual learners spotted the error in the 3D display.
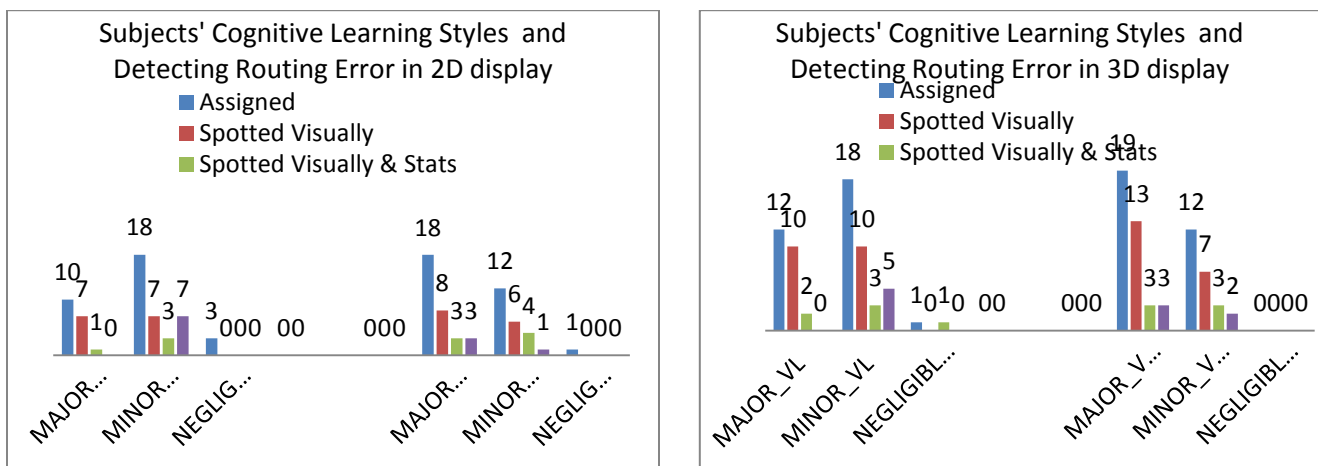

Fig. 12. Subjects cognitive learning styles and performance

- **Study Level**

The results of the tasks performance of subjects from the experiments based on level of education is shown in table 3. Those who did not spot the routing error in the 2D display included participants from all levels of study: PhD OR, PhD computing, undergraduate management, PhD arts and humanities, MSc management, undergraduate management. Therefore there is no obvious connection with particular backgrounds.

- **Experience in Simulation**

On experience in simulation, 40 participants indicated that they had some experience in 2D simulation, while only one participant had some kind of experience in 3D display. Of this number, 19 were randomly assigned to perform the experiment on the 2D display and 21 on the 3D display. Also, only 5 participants (PhD simulation

students) had experience using WITNESS simulation software; 2 of them did the experiment on the 2D display and 3 on 3D display. This indicates that both groups have very similar numbers, which provides further reassurance of similar skill levels between the groups. The result on performance based on experience in simulation is shown in table 3.

- **Technical and Analytical Background of Participants**

We analyzed the performance of subjects based on quantitative and analytical abilities. As indicated in Section 4.3, we used the subject of study or field of specialization of the staff/faculty as indication of participants' technical and analytical abilities. The assignment of subjects based on this factor were: 2, 11, 11, 2, 2, 3 from Simulation PhD, Other Operations Research PhD, Computing and IT, Engineering and Science and Arts, Humanities and Social Sciences for the 2D display respectively, and 3, 10, 12, 1, 1, 4 from the same subjects areas for the 3D group. The results for the error spotting experiments indicate that, participants from both technical and non-technical backgrounds had difficulty spotting the errors in the 2D display. For example, only 50%, 0% and 67%, 67% of participants from the Computing, Information Technology, and Arts, Humanities, Social Science backgrounds respectively spotted the routing and assembly errors (table 3). On the other, participants from all areas of study spotted the routing error on the 3D model and all except one also spotted the assembly error on the 3D model. This indicates that, the performance of participants did not depend on technical and analytical abilities.

Table 3.
Characteristics of the participants

Characteristics Total (n=62)	Detecting Routing Error Experiment 2D n=31			3D n=31			Detecting Assembly Error Experiment 2D n=31			3D n=31		
	S* (%)	MT(RT)**	SD***	S* (%)	MT(RT)**	SD***	S* (%)	MT(RT)**	SD***	S* (%)	MT(RT)**	SD***
Total number	81%	3.04	0.44	100%	1.55	0.22	45%	4.89	1.02	97%	2.05	0.37
Gender												
Male	88%	2.84	0.68	100%	1.61	0.29	50%	4.89	1.54	95%	2.49	0.49
Female	73%	3.31	0.50	100%	1.01	0.24	40%	4.89	1.39	100%	1.01	0.24
Study level												
Undergraduate	82%	1.89	0.35	100%	0.96	0.17	45%	6.57	1.73	100%	1.83	0.60
Masters / postgraduate	80%	4.02	1.25	100%	1.22	0.16	40%	7.98	1.48	100%	2.66	0.71
PhD	80%	3.58	0.71	100%	1.95	0.42	47%	3.67	1.48	93%	1.29	0.26
Staff	-			100%	1.36	0.66	-			100%	4.70	3.99
Current subject area of study												
Simulation PhD	100%	5.03	0.67	100%	1.51	0.76	50%	0.71	-	100%	0.81	0.15
Other Operations Research	91%	2.06	0.32	100%	2.17	0.55	67%	4.95	1.37	100%	1.92	0.79
Other Business and Management	73%	3.44	0.74	100%	1.02	0.14	27%	2.81	2.34	92%	2.61	0.64
Computing, Info Technology	50%	0.85	-	100%	2.57	-	0%	-	-	100%	0.18	-
Engineering, Science	100%	7.46	1.26	100%	0.57	-	50%	9.95	-	100%	3.68	-
Arts, Humanities, Social Science	67%	1.07	0.38	100%	1.65	0.56	67%	8.21	1.60	100%	1.74	0.51
Experience of simulation												
Experience in 2D simulation	89%	3.54	0.54	100%	1.66	0.31	37%	4.89	1.58	100%	2.24	0.51
No Experience in 2D simulation	67%	1.98	0.61	100%	1.34	0.26	58%	4.90	1.43	90%	1.61	0.35

* Successes (percentage of those who spotted the error from each group). ** Mean time (real time in minutes) spent spotting the error. The time spent *does not* include the maximum time of 600 seconds assigned to those who did not spot the error. *** Standard Deviation of the time spent in spotting the error

- **Gender**

The overall performance of subjects on the experiments based on gender as shown in table 3 indicates that there was no significance difference between male (88%) and female (73%), although the male group performed slightly better. However, further diagnostics on how the errors were spotted provided an indication that, more female subjects spotted the routing error on the 2D display visually, while more subjects in the male group used statistics. For example, 56% of female group classified as visual learners spotted the routing error on the 2D display using the animation, with no female visual language learner spotting the same error using the statistics (figure 13). This appears to indicate that, the use of visual display could be more attractive to female users than the male counterpart.
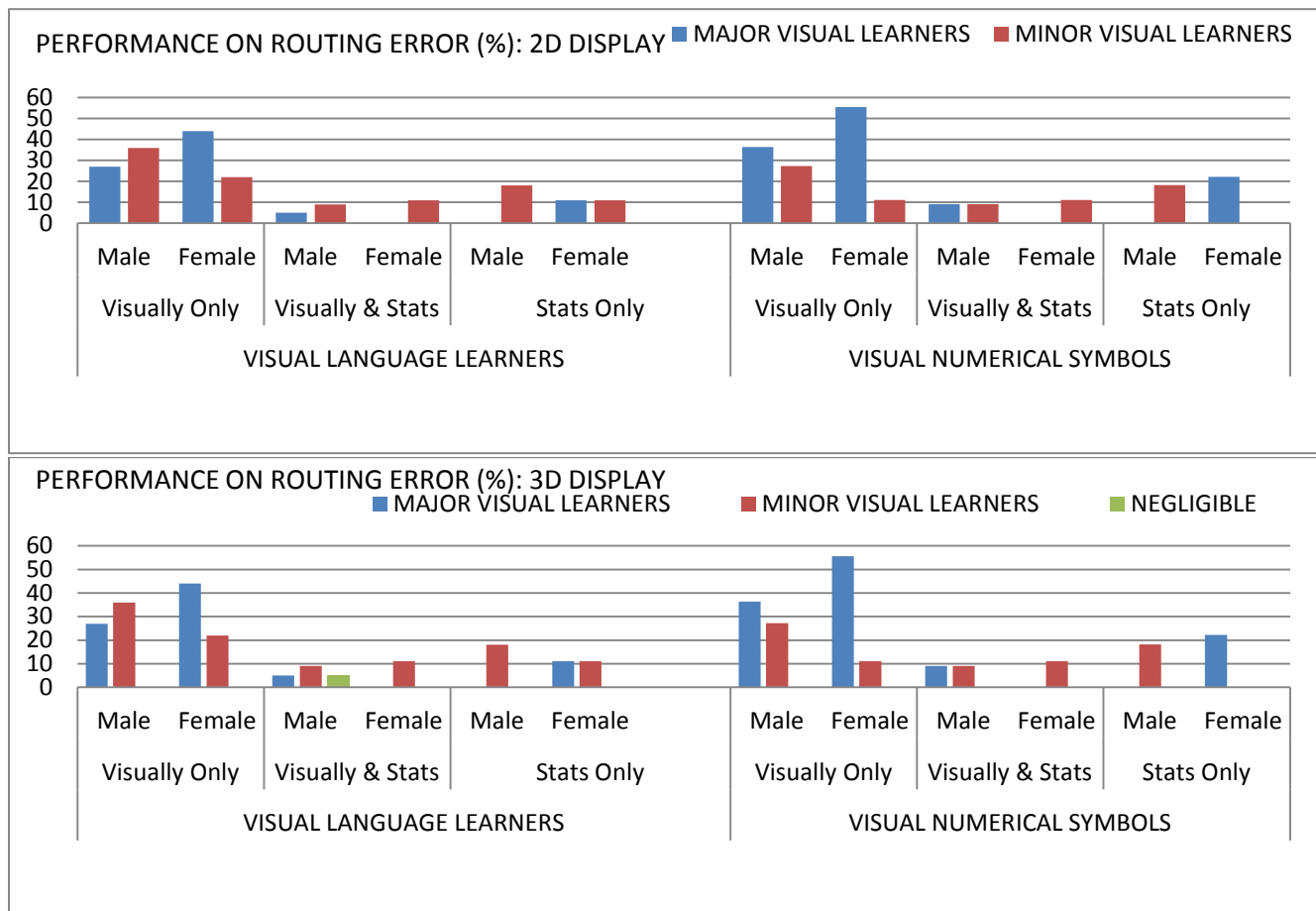

Fig. 13. Performance of Subjects on spotting of routing error based on gender and cognitive learning styles

5.2 **The Impacts of 3D Visualization on Spotting Errors in DES Model (Hypothesis 1)**

We use the results from our experiments on detecting the routing and assembly errors (Section 5.2.1 and 5.2.2) to test the first hypothesis *(H1$_0$ and H1$_1$)*, which examines the effects of the type of display on detecting errors in a DES model.

5.2.1   Detecting Errors – Routing Error

The results for spotting the routing error are shown in Table 4. The performance of the 3D group was much better than the 2D group in spotting the error. Within the 2D group, six out of the 31 participants (19%) were not able to spot the error within the 10 minute time limit of the experiment whereas all 3D perspective participants spotted the error. The average time to spot the error is considerably less for the 3D group being 93 seconds compared to 263 seconds. This means that for this particular experiment changing from the 2D to the 3D perspective display reduced the average time to spot the error by 170 seconds, a reduction of 65%.

Table 4.
Results from the routing error experiment

	2D	3D
Spotted the error	25	31
Did not spot the error	6	0
Average time to spot the error (seconds) *	263.4	93.3
Standard deviation of time to spot the error (seconds) *	204.3	74.4

* A time of 600 seconds (the maximum time in the experiment) was allocated to those who did not spot the error.

In calculating the average time, those who did not spot the error were allocated a time of 600 seconds (the time limit for the experiment). The actual time that they would have needed to spot the error is obviously longer than this and so the average time for the 2D group is slightly underestimated. Therefore, the difference between the groups is also slightly underestimated. However, this is better than ignoring these participants which would have underestimated the average time further. Setting a time limit (although done for practical reasons) also avoids the average being affected by extreme values.

Figure 5 shows the distribution of times to spot the error. There is a big difference in the number of participants who spotted the error within the first minute, being 14 (45%) for the 3D group and only 4 (13%) for the 2D group. All except one participant in the 3D group spotted the error within 194 seconds whereas 14 2D participants took longer than this.

The results show a big difference in performance between the groups. Statistical tests were applied to measure the statistical significance of the results. The exact binomial test was used for the mix of participants who did not spot the error. There were six such participants and all used the 2D display.
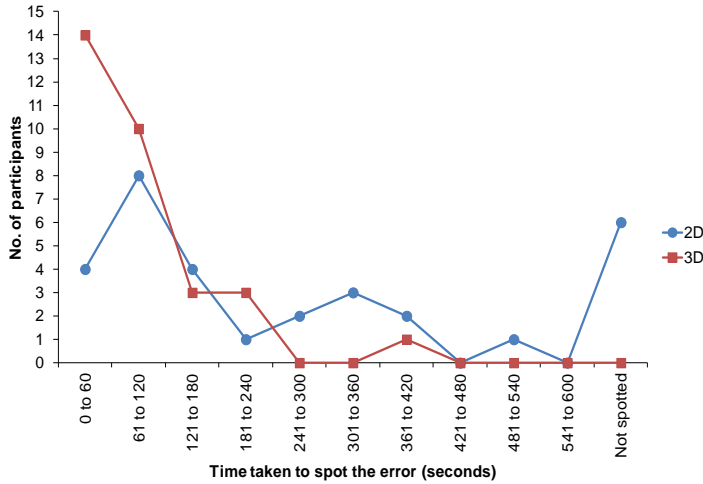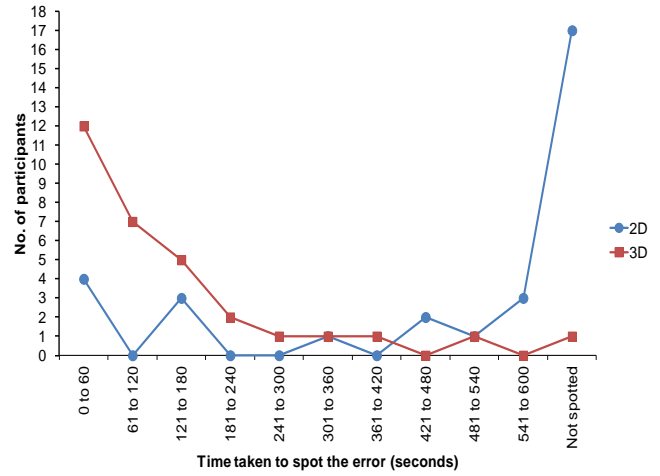
Fig. 14 . Time taken to spot the routing error.

Fig. 15 . Time taken to spot the assembly error.

Using the null hypothesis *(H1_0)* which stated that, the type of display does not have any effect on spotting of error, there would be a probability of 0.5 that a participant who did not spot the error used the 2D display and a probability of 0.5 that they used the 3D display. Since all the six participants used the 2D display, the two tailed exact binomial test is the probability of getting 0 or 6 successes from 6 trials with probability of success 0.5. The probability of this is 0.031 (p=0.031), so the numbers of subjects spotting the error is significant at the 5% level.

On the basis of this result, we reject the null hypothesis *(H1_0)* and accept the alternative hypothesis *(H1_1)* that, 3D Visualization makes it easier to spot errors in DES model.

An exact calculation can also be carried out to compare the numbers that did and did not spot the error. Overall there were six participants who did not spot the error and they all came from one of the groups (the 2D group). There are $^{62}C_6$ combinations of six participants from the overall group of 62. Under the null hypothesis of the type of display having no effect (and ignoring other factors) each combination would be equally likely. The number of combinations that all come from either the 2D group or the 3D group is $2 \times {}^{31}C_6$. Therefore the probability of getting the experiment result is $2 \times {}^{31}C_6 / {}^{62}C_6$ which is 0.024. Therefore, again this indicates that the results for the number of participants spotting the error is significant at the 5% level. This test is better than the chi-square test in this situation (p value of 0.010) both because it is an exact calculation and also because the expected values for not spotted would be small (3 for each category) making the chi-square test value unreliable.

The time taken to spot the error is stronger data than simply whether the error was spotted since it compares the performance of those that spotted the error. A two tailed t-test (assuming unequal variances) on the times for the two groups gives a probability p value of 0.00098. This is therefore very strong evidence of a difference between the groups. The 95% confidence interval for the difference is 91 to 249 seconds. This is also consistent with the above conclusions to reject the null hypothesis *(H1_0)* and accept the alternative hypothesis *(H1_1)* that, 3D Visualization makes it easier to spot errors in DES model.

In the questionnaire after the experiment the participants were asked how they spotted the error. For the 2D participants, 14 spotted it visually, 7 noticed it from the statistics, and 4 used both sources. For the 3D participants, the numbers were 21, 5 and 5 respectively. There is no significant difference between the groups (chi square p value of 0.54) and the percentages for the two groups combined are: visual 63%, statistics 21%, both 16%.

5.2.2   Detecting Errors – Assembly error

The results for spotting the assembly error are shown in Table 5. As with the routing error the performance of the 3D group was much better than the 2D group. In this case 17 of the 2D group did not spot the error (55%) compared to only one of the 3D group (3%). The average time for the 3D group (138 seconds) is 323 seconds less than the 2D group (462 seconds). This is a reduction of 70% which is similar to the reduction of 65% for the routing error. As with the routing results those who did not spot the error were allocated a time of 600 seconds. This applies to many of the 2D group, and so the time to spot the error will be considerably underestimated for the 2D display.

The results indicate that the assembly error was more difficult to spot than the routing error. Compared to the routing results, considerably more of the 2D participants failed to spot the assembly error and the average times for both groups to spot the assembly error are much higher. This is despite the participants being more familiar with the model having tackled the routing error first.

Figure 5 shows the distribution of times to spot the error. From the data, 24 of the 3D group spotted the error within the first 166 seconds compared to only 6 of the 2D group.

Table 5.
Results from the assembly error experiment.

	2D	3D
Spotted the error	14	30
Did not spot the error	17	1
Average time to spot the error (seconds) *	461.5	138.3
Standard deviation of time to spot the error (seconds) *	216.5	147.5

* A time of 600 seconds (the maximum time in the experiment) was allocated to those who did not spot the error.

For this data, there is clearly a very big difference in the numbers spotting the error from each group. Out of the 18 participants who did not spot the error only one was from the 3D group and the two tailed exact binomial test gives a p value of 0.00014. The chi square test can be applied to the numbers who did and did not spot the error, and gives a p value of 7.6×10^{-6}. A two tailed t-test (assuming unequal variances) on the times for the two groups gives probability p value of 7.4×10^{-9}. Both of these very low p values indicate extremely strong evidence of a difference between the groups. The 95% confidence interval for the difference is 229 to 418 seconds. In contrast to the routing error, this error can only be spotted visually as the on screen statistics do not help in this case.

However, this result also supports our earlier conclusions in the routing error (Section 5.2.1). We therefore reject the null hypothesis *(H1$_0$)* and accept the alternative hypothesis *(H1$_1$)* that, 3D Visualization makes it easier to spot errors in DES model.

5.2.3   Detecting Errors – Perceptions of the Model

At the end of the experiment with the car assembly model the participants were asked two questions about their perception of the model regarding the ease of understanding and the helpfulness of the graphics. They were asked to rate both on a 5 point scale and the results are shown in Tables 6 and 7. For both questions most participants in each group gave a rating of 4, and slightly more of the 3D group than the 2D group gave a rating of 5. Therefore the participants generally found the 2D or 3D model easy to understand and found the graphics helpful. Using the chi square test and combining the frequencies of 3 and 4 to avoid low expected values, the differences in results for the two displays are not statistically significant (p values of 0.37 and 0.19 respectively for the two questions). Therefore, the results do not indicate strong evidence of a difference in perceptions.

Table 6.
Responses to the statement "The simulation model was easy to understand".

	Difficult				Easy
	1	2	3	4	5
2D	0	0	5	20	6
3D	0	0	0	22	9

Table 7.
Responses to the statement "The graphics in the simulation helped me to understand the simulation model".

	Not helpful				Helpful
	1	2	3	4	5
2D	0	0	5	17	9
3D	0	0	1	16	14

5.3 The Impacts of 3D Visualization on Model Understanding and Generation of Idea (Hypothesis 2 & 3)

Here, we examine the effects of the type of display on enhancing understanding of a DES model and generation of ideas for decision making. Using these results, we test the two hypotheses *(H2$_0$ and H2$_1$, and H3$_0$ and H3$_1$)* in Sections 5.3.1 and 5.3.2.

5.3.1   Model Understanding

The results for this experiment were obtained by marking the answers provided by the participants after watching the model run for two minutes. Two questions were asked. The first question tested the understanding of the model and was: "Please describe as fully as possible how the system works." We assessed the answers by whether they included the following points:

(1) There are 4 tellers and 3 enquiry staff.

(2) There is one queue for the tellers and one for the enquiry staff.

(3) Customers join one of the queues on arrival.

(4) Some customers join the other queue after being served.

(5) Customers leave the bank after being served by one or both types of staff.

The answers were compared with each of these points and marked as either getting the point correct (right), getting it incorrect (wrong), or not mentioning it (omission). Table VII shows the total in these categories for the 2D and 3D groups (31 participants in each group).

Table 8.
Assessment of the answers to the question "Please describe as fully as possible how the system works" (see text for more details of the points (1) to (5)).

	(1) Staff	(2) Queues	(3) Arrival	(4) 2nd Service	(5) Exit	Total
2D group						
Right	13	11	18	5	6	53
Wrong	0	11	4	4	0	19
Omission	18	9	9	22	25	83
3D group						
Right	8	30	30	23	12	103
Wrong	0	1	0	0	0	1
Omission	23	0	1	8	19	51

For point (1) regarding the number of staff, all participants either got the point correct or omitted it. More of the 2D group got this correct and this is the only point on which the 2D group did better than the 3D group. For the number of queues all except one of the 3D group got this correct whereas there were many incorrect answers in the 2D group. Similarly all except one of the 3D group mentioned the customers joining the queues on arrival whereas there were several omissions and a few incorrect answers in the 2D group. Most of the 2D group did not mention points (4) and (5). We hypothesize that the omissions for (4) are likely due to it going noticed whereas omissions for (5) are probably because it was not thought worth mentioning in the answer. Both points were included by more of the 3D group. Overall the 3D group did better with about twice as many correct points.

Scores for each participant were calculated by awarding one mark for a correct point, zero marks for an omission, and minus one for an incorrect point. Therefore the maximum mark is five. Figure 7 shows the distribution of marks across the 31 participants in each group. The averages are 1.10 for the 2D group and 3.29 for the 3D group. A t-test on the scores gives a p value of 4.1×10^{-7} (p<0.05) and so there is very strong evidence of a difference. As shown in Figure 9, 28 of the 3D group (90%) scored three or more in contrast to only nine of the 2D group (29%). The display perspective has a surprisingly large impact on the ability of the participants to recall and describe the system.

Our null hypothesis $(H2_0)$ stated that, 3D visualization does not result in a better understanding of the model than the 2D display. As this is not consistent with our results, we reject the null hypothesis $(H2_0)$ and accept the alternative hypothesis $(H2_1)$ that, 3D Visualization does enhance spotting of errors in a DES model.

5.3.2 Generation of Ideas for Decision Making

The second question that the participants answered was about decision making and how to improve the system. The question was: "The bank wishes to provide good service without incurring excessive cost, what changes to the system would you recommend and why?". The answers were assessed as to whether they included the following two points:

(1) The tellers are busy most of the time and long queues build up for the tellers. Therefore the number of tellers should be increased.

(2) The enquiry staff are idle for most of the time and so several enquiry staff could be re-deployed as tellers.

As shown in Table 9, each of the points was correctly stated by most participants in both groups, with no wrong answers. The results are therefore very similar for the two groups with average marks of 1.71 and 1.68 (out of a maximum of 2) for the 2D and 3D groups. There is no statistically significant difference between the groups (t-test p value of 0.85). The marks of the individual participants were: 0 marks: 2 2D, 4 3D; 1 mark: 5 2D, 2 3D, 2 marks: 24 2D, 25 3D. The display type clearly has had a negligible impact on the ability to answer this question.

The results support the null hypothesis $(H3_0)$ which stated that, 3D visualization does not enhance generation of ideas for improving business decisions (Section 3.2). This is despite the previous question indicating that the display made a big difference to the understanding and recall of the way the system worked. It could be that identifying ideas for improving the system does not depend much on the display or on the detailed understanding of the system, at least for this model. However, since most participants got this question completely correct it may also be that this question was too easy to show up a difference between the displays.
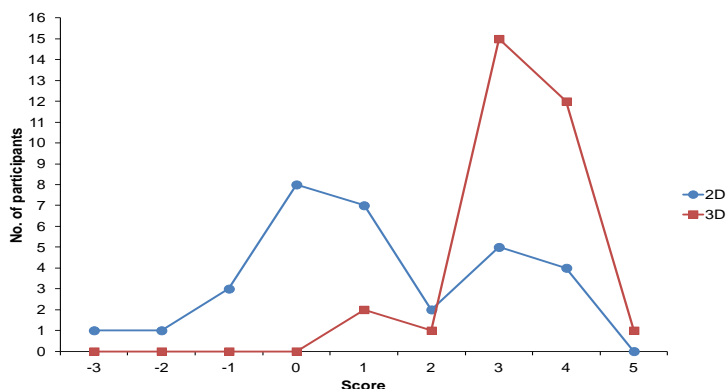

Fig. 16 . Scores for the answers for the description of the system.

Table 9.
Assessment of the answers to the question "The bank wishes to provide good service without incurring excessive cost, what changes to the system would you recommend and why?" (see text for more details of the points (1) and (2)).

	(1) Tellers	(2) Enquiry staff	Total
2D group			
Right	29	24	53
Wrong	0	0	0
Omission	2	7	9
3D group			
Right	27	25	52
Wrong	0	0	0
Omission	4	6	10

5.3.3   Model Understanding – Perceptions of the Model

After the model understanding experiment the participants were asked the same two questions about their perceptions of the bank model that were used for the car assembly model (section 3.3). As for the car assembly model the results were similar for the 2D and 3D groups with most participants giving a rating of 4 or 5 on the 5 points scale. For "the simulation model was easy to understand" the frequencies for ratings 1 to 5 were: 2D group: 0, 0, 0, 18, 13; 3D group: 0, 0, 0, 20 11. For "the graphics in the simulation helped me to understand the simulation model" the frequencies for ratings 1 to 5 were: 2D group: 0, 0, 1, 19, 11; 3D group: 0, 0, 0, 19, 12. Therefore most participants gave a high rating for both criteria for the model they were using and there is no significant evidence for a difference in ratings for the two displays (chi square test p values of 0.60 and 0.79 respectively).

5.4 The Impacts of 3D Visualization on Model Acceptability (Hypothesis 4)

In this experiment, we examined the effects of the type of display on user's preference and acceptance of the model as accurate representation of the real system (the imaginary car assembly operation in this case). We tested the fourth hypothesis *(H4$_0$ and H4$_1$ developed in Section 3.3)* based on the results if this experiment.

This experiment was based on subjective model assessment. It only took place if there was enough time left within the 30 minutes allocated. Therefore the number of participants was less than for the previous experiments with 47 out of the original 62 taking part. The participants watched the car assembly model being run for two minutes each for the 2D and 3D displays, starting with the display they didn't use on the detecting errors experiment. They were then asked which display they considered a more accurate representation of the imaginary system as contained in the description handed to them at the start of the experiment. The results were that 3 (6%) chose 2D, 41 (87%) chose 3D, 3 (6%) were indifferent. They were also asked which display they considered easier to understand in terms of the processes or interactions between the model elements. The results for this question were very similar: 1 (2%) chose 2D, 42 (89%) chose 3D, 4 (9%) were indifferent. The results show a very strong subjective preference (acceptability) for the 3D display amongst the participants. The chi square test comparing those who preferred 2D or 3D with expected values of equal numbers gives p values of 1.0 $\times 10^{-8}$ and 4.0×10^{-10} respectively.

On the basis of these results, we reject the null hypothesis $(H4_0)$ that, 3D visualization does not improve model acceptability than 2D display (Section 3.3), and accept the alternative hypothesis $(H4_1)$ that, 3D visualization does improve acceptability of the model.

6. The Scope, Implications and Limitations of Study

6.1 The Scope

The scope of this study is determined by the DES tasks performed in the experiments, the objectives of each experiment and the decision variables, and the nature of the business operations. The first experiment involves validation and verification of model of a car assembly operation. Providing participants with a model description and being asked to find any errors that contrast the given description depicts a verification test, and detecting errors in the model against set parameters involves model validation as in [6,20]. The application area was based on a factory operation and would be relevant for any objective about factory operations such as improving throughput, scheduling, minimize work in progress, and optimal factory layout.

In the second experiment, we asked the subjects to recommend ways of improving service operation (providing better customer service without excessive costs). Although the simulation model was based on a commercial bank operation, the results would be relevant to any objective about customer service improvement and efficient utilization and assignment of human resources in any organization.

The third experiment involved participants' subjective preference of the DES model as true representation of the actual system, implying user's perception of credibility, and the confidence reposed on the model to enhance acceptability and usability [31].

The study therefore has implications for a very wide range of problem types and domains. In most simulation decision support applications it will be beneficial for both the modeler and the client to understand the model display more easily. As discussed in Section 2, this will help in various tasks such as building the model, verification and validation, and experimentation.

However, the importance of certain aspects of the display will vary depending on the objectives and the decision variables. For example, where the objective is to improve factory performance by optimizing the layout, getting good accuracy in the relative positions of each element will be important. If the focus is on the use of resources to reduce queues in a service process, then the display needs to show clearly (amongst other things) the queue sizes and the tasks being carried out by the resources. Nevertheless, if a 3D perspective view makes the model display easier to understand and to relate to the real system then this would be an advantage in both cases. The nature of the system may perhaps limit the benefits if, for example, the system is very large and complex and therefore difficult to visualize. The two examples in the experiments were quite small systems. If there are a very large number of elements then a 3D perspective might make it harder to see all the elements, although features such as the ability to zoom in and move to different areas easily may overcome this.

6.2 Implications of the Study

The results of the experiments and the conclusions provide convincing evidence that, 3D visualization can bring several added benefits to DES modeling within the application domain as defined in Section 6.1. This include enhancing model validation and verification, understanding, credibility and acceptability, and usability, which can lead to the overall success of the simulation project [30,40]. Thus, despite any skepticisms by some researchers and industry practitioners especially those in academia, 3D modeling and simulation in DES appears to be an acceptable next step in advancing DES practice.

Notwithstanding the above benefits of 3D visualization in DES, there are possible dangers of using 3D visualization as a quick-fix to ensuring model credibility and acceptability [40]. This indicates the possibility of users of project owners easily accepting the DES as a decision support tool based on a pretty interface without considering its technical soundness.

6.3 Limitations of the Study

The use of students as participants in the experiments with little or no experience in discrete event modeling and simulation, especially in 3D modeling meant that, the subjects were limited to just observing the 2D and 3D animations on the 2D display and 3D perspective view. This therefore limited the activities that could be performed in this study. For example, tasks such as model development, model testing and experimentation could not be undertaken. However, this does not rule out the validity of the interesting results and implications for DES modeling. Also, both the experimental methods and the use of students as subjects in the experiments is consistent with previous experiment that investigated the efficacy of 2D display against non-visual interface [11].

7. Summary and Conclusions

The experiments showed that making a relatively small change to the visual display in altering the perspective from an overhead 2D plan view to a 3D perspective view resulted in much better performance in spotting errors and understanding the model. The average time to spot the errors in the car assembly model was reduced by 65% and 70% for the two errors. The participants also performed much better with the 3D display than the 2D display in being able to recall and describe the bank model. There was very little difference between the displays in the ability of the participants to suggest improvements to the system although this was probably at least partly because this question was too easy. Almost all participants preferred the 3D display when asked to compare the two displays.

The visual display is very important in simulation in enabling both the modeler and the client to observe the behavior of the model. It can affect many aspects of the simulation project particularly verification, validation, the understanding of the system, and the generation of ideas for system improvement and for experimentation. It also plays a major role in facilitating interaction between the modeler and the client.

We were surprised by the extent of the difference in performance caused by such a small change in the display. This indicates that the detailed way that the visual display is designed may have a major effect on the outcome of several of the simulation tasks and therefore on whether the project is successful. One implication is that the design of the visual display should receive more emphasis in simulation software, simulation teaching and simulation practice.

More research in this area would help to extend and generalize the results. Such research could include doing similar experiments but with different models, investigating other aspects of the display, and analyzing the process of interaction between the user and the display.

Acknowledgments

References

[1] I. J. Akpan, R. J. Brooks, Users' perceptions of the relative costs and benefits of 2D and 3D visual displays in discrete-event simulation, Simulation 88(4) (2012) 464-480.

[2] I. J. Akpan, Empirical Study of the Impacts of Virtual Reality on Computer Simulation. PhD Thesis, Lancaster University, UK (2006).

[3] M. Alemparte, D. Chheda, D. Seeley, W. Walker, Interacting with discrete simulation using on line graphic animation, Computers & Graphics, 1(4) (1975) 309–318

[4] D. F. Andersen, J. A. M. Vennix, G. P. Richardson, E. A. J. A. Rouwette, Group model building: Problem structuring, policy simulation and decision support, J Opl Res Soc., 58(5) (2007) 691-694.

[5] G. Au, R. J. Paul, Visual interactive modelling: A pictorial simulation specification system, European Journal of Operational Research, 91(1) (1996) 14–26

[6] O. Balci, Validation, verification, and testing techniques throughout the life cycle of a simulation study. Annals of operations research, 53(1) (1994) 121-173.

[7] M. Barnes, An introduction to Quest, in: Proceedings of the Winter Simulation Conference, (Eds.) S. Andradottir, K. Healy, D. Withers, B. Nelson, (1995) 432-436.

[8] P. C. Bell, Visual interactive modelling: The past, the present, and the prospects, European Journal of Operational Research, 54(3) (1991) 274-286.

[9] P. C. Bell, C. K. Anderson, D. S. Staples, M. Elder, Decision-makers' perceptions of the value and impact of visual interactive modelling. Omega, 27(2) (1999) 155-165.

[10] P. C. Bell, R. M. O'keefe, Visual interactive simulation — history, recent developments, and major issues. Simulation, 49(3) (1987) 109-116.

[11] P. C. Bell, R. M. O'keefe, An experimental investigation into the efficacy of visual interactive simulation. Management Science, 41(6) (1995) 1018-1038.

[12] V Belton, M. D. Elder, Decision support systems: Learning from visual interactive modelling, Decision Support Systems, 12(4–5) (1994) 355-364.

[13] V. Čerić, Visual interactive modeling and simulation as a decision support in railway transport logistic

operations. Mathematics and computers in simulation 44(3) (1997) 251-261.

[14] J. G. Crookes, Simulation in 1981. European Journal of Operations Research, 9(1) (1982) 1-7.

[15] D. A. Guttentag, Virtual reality: Applications and implications for tourism. Tourism Management, 31(5) (2010) 637-651.

[16] J. Hugan, QUEST – Queuing event simulation tool. Proceedings of the Winter Simulation Conference, (1995) 432-436.

[17] R. D. Hurrion, European, Visual interactive modeling, Journal of Operational Research 23 (1986) 281-287

[18] R. D. Hurrion, Implementation of a visual interactive consensus decision support system, European Journal of Operational Research, 20(2) (1985) 138-144.

[19] R. D. Hurrion, R. J. R. Secker, Visual interactive simulation an aid to decision making Omega, 6(5) (1978) 419-426.

[20] V. R. Kamat, J. C. Martinez, Validating complex construction simulation models using 3D visualization, Systems Analysis Modelling Simulation, 43(4) (2003) 455-467.

[21] V. R. Kamat, J. C. Martinez, Enabling smooth and scalable dynamic 3D visualization of discrete-event construction simulations, in: Proceedings of the 2000 Winter Simulation Conference, Arlington, USA, 2 (2001) 1523-1533.

[22] W. L. Koh, S. Zhou, Modeling and simulation of pedestrian behaviors in crowded places. ACM Transaction on Modeling and Computer Simulation, 21(3) (2011) Article 20, 23 pages.

[23] J. Kuljis, HCI and simulation packages. In Proceedings of the 28th conference on Winter simulation, IEEE Computer Society, (1996) 687-694.

[24] J. Kuljis, R. J. Paul, C. Chen, Visualization and simulation: Two sides of the same coin? Simulation 77(3-4) (2001) 141-152.

[25] J. A. List, S. Sadoff, M. Wagner, So you want to run an experiment, now what? Some Simple Rules of Thumb for Optimal Experimental Design. Experimental Economics, 14(4) (2011) 439-457.

[26] A. M. Law, Simulation Modeling and Analysis, 4th edition, McGraw Hill, New York, 2007.

[27] T. S. Mujber, T. Szecsi, M. S. J. Hashmi, Virtual reality applications in manufacturing process simulation, Journal of Materials Processing Technology, 2004; 155-156: 1834-1838

[28] S. I. Park, G. Lee, M Kim, Do students benefit equally from interactive computer simulations regardless of prior knowledge levels? Computers & Education, 52(3) (2009) 649-655

[29] J. Quarles, P. Fishwick, S. Lampotang, I. Fischler, B. Lok, A mixed reality approach for interactively blending dynamic models with corresponding physical phenomena, ACM Transaction on Modeling and Computer Simulation, 20(4) (2010) 23 pages.

[30] S. Robinson, Discrete-event simulation: from the pioneers to the present, what next? Journal of the Operational Research Society, 56(6) (2005) 619–629.

[31] M. Rohrer, Seeing is believing: the importance of visualization in manufacturing simulation, in J Joines, R. Barton, K. Kang and P. Fishwick (Eds.), Proceeding of the 2000 Winter Simulation Conference, IEEE Computer (2000) 211-1216,

[32] J. M. Rosen, H. Soltanian, R. J. Redett, D. R. Laub, Evolution of virtual reality [Medicine], Engineering in Medicine and Biology Magazine, IEEE,15(2), (1996) 16-22.

[33] M. M. Sebrechts, J. V. Cugini, S. J. Laskowski, J. Vasilakis, M. S. Miller, Visualization of search results: a comparative evaluation of text, 2D, and 3D interfaces, in Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, (1999) (3-10).

[34] J. P. Shim, M. Warkentin, J. F. Courtney, D. J. Power, R. Sharda, & C. Carlsson, Past, present, and future of decision support technology, Decision support systems, 33(2) (2002) 111-126.

[35] C. B. Tilanus. Failures and successes of quantitative methods in management, European Journal of

Operational Research, 19(2) (1985) 170-175.

[36] M. Tory, A. E. Kirkpatrick, M. S. Atkins, and T. Möller, Visualization Task Performance with 2D, 3D, and Combination Displays, Visualization and Computer Graphics, IEEE Transactions on, 12(1) (2006) 2-13.

[37] E. C. Valentin, R. A. Bijlsma, & V. de Gast, Empowering decision support with simulation technology-Scenario Navigator, in: Winter Simulation Conference, IEEE Computer Society, (2008) 236-244.

[38] D. J. Van Der Zee & J. G. Van Der Vorst, A modeling framework for supply chain simulation: Opportunities for improved decision making. Decision Sciences, (2005). 36(1), 65-95.

[40] A. P. Waller, J. Ladbrook, Experiencing virtual factories of the future, in: Proceedings of the Winter Simulation Conference, (2002) 513-517.

[41] Z. Wang, C. K. Chui, Y. Cai, C. H. Ang, S. H. Teoh, Dynamic linear level octree-based volume rendering methods for interactive microsurgical simulation. International Journal of Image and Graphics, 6(02), 155-171.

[42] S. Wenzel, U. Jessen, The integration of 3-D visualization into the simulation-based planning process of logistics systems, Simulation 77(3-4) (2001) 114-127.

[43] D. M. Wilhelm, K. Ogan, C. G. Roehrborn, J. A. Cadeddu, M. S. Pearle, Assessment of basic endoscopic performance using a virtual reality simulator. Journal of the American College of Surgeons, 195(5) (2002) 675-681.

[44] C. C. Wu, N. B. Dale, L. J. Bethel, Conceptual models and cognitive learning styles in teaching recursion. In ACM SIGCSE Bulletin 30(1) (0000) 292-296.

[45] T. C. Yang, G. J. Hwang, S. Jen-Hwa Yang, Development of an Adaptive Learning System with Multiple Perspectives based on Students' Learning Styles and Cognitive Styles. Journal of Educational Technology & Society, 16(4) (2013).

[46] Y. Zhong, B. Shirinzadeh, Virtual factory for manufacturing process visualization. Complexity International, 12, (2008) 1-22.

[47] Z. Zhou, A. D. Cheok, T. Chan, J. H. Pan, Y. Li, Interactive entertainment systems using tangible cubes, in Proceedings of the First Australian Workshop on Interactive Entertainment (2004) (19).

[48] M. S. Zywno, A contribution to validation of score meaning for Felder-Soloman's index of learning styles, in: Proceedings of the 2003 American Society for Engineering Education Annual Conference & Exposition (2003) 119(1-5).