Stress training and simulator complexity: why sometimes more is less

Jennifer G. Tichon *, Guy M. Wallis *

* The School of Human Movement Studies, The University of Queensland, St Lucia, Australia

First published on: 30 November 2009

To cite this Article Tichon, Jennifer G. and Wallis, Guy M.(2009) 'Stress training and simulator complexity: why sometimes more is less', Behaviour & Information Technology,, First published on: 30 November 2009 (iFirst)

To link to this Article: DOI: 10.1080/01449290903420184
URL: http://dx.doi.org/10.1080/01449290903420184
Stress training and simulator complexity: why sometimes more is less

Jennifer G. Tichon* and Guy M. Wallis

The School of Human Movement Studies, The University of Queensland, St Lucia, 4072, Australia

(Received 20 January 2009; final version received 15 October 2009)

Through repeated practice under conditions similar to those in real-world settings, simulator training prepares an individual to maintain effective performance under stressful work conditions. Interfaces offering high fidelity and immersion can more closely reproduce real-world experiences and are generally believed to result in better learning outcomes. However, absolute fidelity in stress training is not critical for skills to be transferable. The present study compared the performance outcomes achieved by trainees using two different simulator types to complete a training program aimed at improving decision-making skills. The purpose of this research was to assess both the overall level of training effectiveness and to determine whether performance levels were influenced when high (160 degree curved wide screen) versus low fidelity (small cab-based flat screen) simulator types were in use. Sixty-three train drivers drove for 40 min on a simulated track on which they encountered four major high stress driving events. One year later, 42 of the original drivers returned and repeated the training scenario a second and third time. Results revealed trainees using the lower fidelity flat screen simulator made fewer errors in both years than trainees using the high fidelity curved screen simulator. The implications of these results are discussed.

Keywords: task analysis; simulator; stress; cognitive training; presence

1. Introduction

In work environments where heavy, human operated machinery is at work, accidents are an unfortunate but often inevitable part of everyday life. An essential element, in training to respond to workplace accidents effectively, is how to train staff to cope with ill-defined problems in the face of high stress environments constrained by factors such as insufficient or unreliable information, time constraints and or conditions of threat of physical injury (Salas et al. 1999). Emergency response skills, particularly those that degrade under stress such as critical thinking and decision-making, have been highlighted across a number of industries as in need of attention.

The question of how to most effectively train higher-order mental abilities in these situations is now emerging as a focus of research interest. Suitable training in critical thinking under stress can hugely reduce the impact of workplace accidents, not only in terms of the immediate physical well-being of staff and clients, but also the long-term mental health of those involved. Simulators through recreating real-world situations in virtual environments provide a compelling opportunity for meeting the need for high affect training. Accordingly simulator training has been established as a core component of safety training programs. High fidelity, interactive simulation is typically achieved through complete immersion in brief, stressful and complex scenarios.

2. Stress exposure training via simulation

When training goes beyond the acquisition of standard, required knowledge and abilities and instead aims to prepare trainees to perform effectively in a stressful environment, it is commonly referred to as stress exposure training (Driskell and Johnston 1998). Stress exposure training focuses on developing those cognitive skills required to maintain effective performance under stress. The overall goal of training via simulation is to build confidence in staff in their own ability to perform under adverse conditions (Stetz et al. 2006). Sub-goals of stress exposure training include gaining specific knowledge of and familiarity with the operational environment to assist trainees to form accurate expectations of the environment, increasing their ability to predict outcomes, avoid errors and decrease their propensity to be distracted by novel sensations (Driskell and Johnston 1998, Hulse and Memon 2006). Simulator training has been shown to be effective when the trainee experiences success or a sense of mastery during the training (Maschuw et al. 2008). This confidence translates into being less aroused physiologically, less distracted and more likely to focus on the task.

*Corresponding author. Email: jtichon@uq.edu.au
Simulators are used for instruction and training in areas such as commercial and military aviation, battlefield management, building construction, and first responder agencies because they provide a safe alternative to replicating high-risk real-world training scenarios. Real-world renditions of these situations are too costly and dangerous to provide opportunities for staff to repeatedly practise responding to high stress workplace situations (Tichon et al. 2003). It is the ability of immersive simulation to engender the same psychological and psychological responses experienced in the real world that has been seen as valuable (Romano and Brna 2001).

3. High fidelity or low fidelity?
A crucial element of the approach is to provide the opportunity for repeated practice under operational conditions similar to those likely to be encountered in real-world settings. The aim is to incite the same emotional (stress) reaction and stems from research demonstrating that for some tasks normal training procedures did not improve task performance when the task was later performed under stress conditions (Zakay and Wooler 1984).

The use of simulator technology alone, however, cannot ensure a successful training outcome without prior consideration of the best use of its features (Cannon-Bowers and Salas 1998, Wallis et al. 2007). Prior research has described a number of variables of interest to simulation-based critical decision-making training but the best use of these is not always agreed on (Thomas 2003). There is a debate over the level of fidelity required for successful training outcomes. Some suggest that when using simulation for cognitive skills training, the fidelity should only match the requirements of the training such as facilitating reflection and learning in the training experience (Hoffman et al. 2001). Yet, it is also recognised that the similarity of the training environment to the actual conditions under which the trainee will perform in the real world is an important factor in simulation training design. While realistic simulations do not always translate directly to training effectiveness, good outcomes have been reported when simulations have reproduced realistic tasks and afforded trainees an engaging experience they can relate to their real world (Baker et al. 2005). A study of simulator training in underground cave structures reported task performance was two to three times faster in the high-immersion condition, and question responses were 3 to 10 times more accurate (Bowman and McMahon 2007).

Despite increasing affordability high-end immersive technologies such as head mounted displays and multi-screen stereoscopic projection displays remain too expensive for widespread use. In training applications significant improvements in performance using simulators in comparison to alternative training support tools would have to be proven to justify the expense. Consequently, questions such as ‘How much immersion is enough?’ and ‘When are driving environments real enough?’ are being more and more frequently asked (Walshe et al. 2005, Bowman and McMahon 2007). Research has suggested that in driving applications specifically difficulties can arise if users are not immersed sufficiently in simulated driving scenes (Nanyue et al. 2006). Is there a compromise between the traditional alternatives of textbooks and classroom teaching and the high-end, high fidelity that can still produce a level of realism sufficient to gain the required performance outcomes of stress exposure training?

The overall purpose of this research was to assess the level of training effectiveness in terms of enhanced decision-making under stress. To achieve this, comparisons among performance outcomes over three different training exposures was made. Additionally, the study design allowed for comparison of the performance results of trainees using a high fidelity curved wide-screen simulator with the performance of trainees using a smaller cab-based flat screen simulator. Reactions to four separate stressful events were examined for effect of simulator type on learning outcomes.

As it is the illusion of being in a different place that is essential for the effectiveness of these training applications (Romano and Brna 2001, Baker et al. 2005), the two simulator types were also compared in terms of their effectiveness in generating immersion and thereby engaging the driver’s attention. Presence is a variable widely accepted as being positively related to enhanced learning and performance in immersive training environments and therefore highly desirable (Witmer and Singer 1998). It facilitates the environment replicated in the simulator being interpreted as real by the trainee thereby making it more likely decision-making skills developed will be transferred to the real world. Presence has been described ‘as the subjective experience of being in one place or environment, even when one is physically situated in another’ (Witmer and Singer 1998). Research investigating the role of presence in facilitating the effectiveness of simulator-based educational training indicates the two seem to be inextricably tied.

4. Method
4.1. Participants
4.1.1. Year one
In 2006, the first year of the project, 63 participants volunteered to take part in the study. The 61 males and
2 females possessed an average of 14.8 (SD = 9.68) years’ experience driving trains, with a range between 1 and 30 years of experience. The drivers were aged from 29 to 58 years of age with a mean age of 43.3 (SD = 8.14) years.

Twelve participants were trained in the wide-screen reality centre while 51 were trained in the cab-based simulator. The unequal sample size for the simulators was due to drivers being given the opportunity to self-select the simulator they would use for training. This training was part of their annual training and drivers were allowed to choose their simulator as per normal operating procedures. As drivers were being assessed via their simulation performance they were allowed to select the simulator they felt the most comfortable and confident using. Self-selection was also allowed for the study as it did not result in a driver being randomly assigned a simulator with which they were unfamiliar. A no-control group design was used at the request of the rail company which required all drivers to undergo the annual safety training.

4.1.2. Year two

Forty-two participants who participated in the 2006 study above returned in 2007. However, due to a malfunction with the wide-screen simulator eight of these drivers were reassigned the cab-based simulator. As this was not the simulator they had used in 2006 these eight sets of results could not be used in this study. Because of the expense of bringing active drivers off-line for the research it could not be rescheduled for another time. Of the remaining 34 participants, 33 were males and 1 was a female. They possessed an average of 12.9 (SD = 9.94) years’ experience driving trains, with a range between 1 and 32 years of experience. The drivers were aged from 30 to 59 years of age with a mean age of 43.5 (SD = 8.66) years. Again participants volunteered to take part in the study.

4.2. Equipment

Results were obtained from two simulator types:

(1) Cab-based. A reconfigurable, full-size, in-cab simulator which can reproduce the arrangement of standard in-cab train controls. Visual display input was provided via a flat projection screen (3 m wide × 2 m high) and driven by a single-channel, SGI Onyx 2 (Four 400 MHz CPUs).

(2) Wide-screen. A curved screen reality centre 7.3 m in diameter offering a 160 degree field-of-view from a centrally located train control desk. The desk has force-feedback controls but no enclosure. The system is driven by a three-channel SGI Onyx 3000 (Twenty-four 500 MHz CPUs).

4.3. Training scenario

The training scenario involved entering a worksite operating under Track Work Authority (TWA), encountering detonators and an outer hand signaler at the bottom of the flyover. Drivers then had to stop at a station and remember prior warnings when departing the station before encountering further detonators, an inner hand signaler and finally the worksite. A second incident involved a Condition Affecting the Network (CAN). Specifically, the CAN was a failed level crossing attended by hand signalers. After receiving advice of the failed crossing the driver next encounters school children trespassing on the track, followed by a station and finally a tunnel. These interleaved events were designed to interfere with the driver’s concentration potentially resulting in the driver forgetting to prepare for the failed crossing. Finally, drivers were required to pass a failed absolute signal at stop immediately before a station located in a tunnel. Each of these events were further complicated through increasing the driver’s workload pressures by rapidly changing operational conditions such as variations to track infrastructure including curvature, grade, speed and signal siting and sequence. The rate at which these conditions were changed often resulted in a driver feeling under intense time pressure. Incidents also included the receipt of information into the driving cab via multiple sources, both visual and auditory.

4.4. Procedures

To determine whether stress training via simulation was improving performance the decision making skills of 63 drivers undertaking their annual simulator training were recorded in 2006. Rail incidents encountered during the simulator training exposure were analysed to compile a checklist of correct actions the driver must perform simultaneously for each key event in addition to a checklist of possible errors. From the 4 major incidents encountered in the training scenario outlined above, 36 correct actions were required to maintain safe and effective train travel on the network and 45 possible errors were identified. All drivers attended a classroom refresher course on rules and regulations and undertook a 5 min practice run on the simulator before their simulation training. Trainers completed the checklist of correct actions and possible errors while drivers undertook the 40 min simulator training run.
In 2007, 34 of the drivers who had participated in 2006 returned for training to undertake the same simulated driving scene using the same simulator type they had been tested on the previous year. This is referred to as Trial 1. To test the premise that repeated exposure will inoculate drivers and result in enhanced decision-making skills, the drivers were also required to undertake the driving scene after a short half hour time delay. This is referred to as Trial 2. Comparing across exposures, Trial 1 against Trial 2, decision making performance could then be assessed after a short time delay (on the same day) and after a long time delay (12 month interval). All the same performance data were gathered during the first and second sessions as was collected in 2006 and drivers used the same simulator type which they had used in the first instance.

In 2007 in between the two simulator exposures the drivers completed two surveys the Presence Questionnaire (Witmer et al. 2005) and the IGroup Presence Questionnaire (Schubert et al. 2001). The surveys were selected to measure the concept of presence in two ways, through introspective self-report and causal factor self-report measures to gain data on the effectiveness of both the wide-screen and the cab-based simulators in generating immersion. The IPQ measures spatial presence by asking questions that require participants report on their own individual experience. The PQ explores four factors thought to underlie presence: involvement, sensory fidelity, adaptation and interface quality.

5. Results

Results were converted to an absolute percentage error score allowing scores across simulator types to be compared as seen in Table 1.

5.1. Repeated practice

The comparison between the 2006 and 2007 Trial 1 results indicated that total driver error rates were higher in 2007 (11.7%) than 2006 (5.2%). However, in 2007, the opportunity for repeated practice resulted in improved error rates. A 2 (simulator) \(\times\) 4 (decision point) \(\times\) 2 (trial) mixed ANOVA was performed on the 2007 survey data and revealed two significant main effects, one for trial (\(F = 13.84; df = 1, 31, p = 0.001\)) and the other for decision points (\(F = 17.29; df = 1.31, p = 0.000\)). The main effect of trial revealed that the error rate was significantly greater during the driver’s first exposure (Trial 1) than during their second exposure (Trial 2) (11.7% vs. 2.6%, respectively). Therefore, when given the opportunity for a second exposure in 2007 driver’s error rates decreased significantly.

5.2. Error rates by simulator type in 2006

The total error rates by simulator were then compared to determine whether the choice of simulator altered the effectiveness of the training.

In 2006, 12 participants were trained on the wide-screen while 51 were trained on the smaller cab-based simulator. A 2 (simulator) \(\times\) 4 (decision points) mixed ANOVA was performed and revealed to be significant (\(F = 55.39, df = 1.61, p = 0.018\)) for the main effect of simulator. This result indicates that participants had greater error rates when using the 160 degree curved wide-screen (9%) than the flat cab-based screen (4%) over the four major decision points investigated. This result was interesting in that the effect of simulator held over all sections of the virtual track. At each point in the scene when decisions had to be made to avert accidents or other incidences drivers using the wide-screen consistently made more errors than their colleagues using the cab-based screen as seen in Table 2. The effect of simulator was not influenced by the type of problem being negotiated on the virtual track. No significant effect was found for the main effect of decision points.

5.3. Individual error types

In terms of specific errors, in 2006, the highest error rates were all recorded in the wide-screen simulator. The highest number of errors was recorded (58%) when drivers were required to report the Children on

<table>
<thead>
<tr>
<th></th>
<th>2006 (CS (n = 12))</th>
<th>2007 (Trial 1) (CS (n = 4))</th>
<th>2007 (Trial 2) (CS (n = 4))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average error rate</td>
<td>3.4</td>
<td>4.75</td>
<td>2.25</td>
</tr>
<tr>
<td>Error rate as % of total number of actions</td>
<td>9.5</td>
<td>13.2</td>
<td>6.25</td>
</tr>
<tr>
<td>Total % (simulator totals combined)</td>
<td>5.2</td>
<td>11.7</td>
<td>2.6</td>
</tr>
</tbody>
</table>

CS, curved screen; FS, flat screen.
Track incident to the guard after the signaller had been advised via radio. The second highest error rate (50%) was recorded again in the wide-screen simulator during the absolute signal at stop incident, during which the driver is required to either advise the signaller to standby or wait to answer the radio after the train is stopped. The third highest error rate, again recorded using the wide-screen, occurred in the Failed Level Crossing incident (42%).

5.4. Error rates by simulator type in 2007

Because of a simulator malfunction affecting some drivers, in 2007 it was not possible to test the significance of differences between the two simulators in that year. Nonetheless, the general trend was consistent with the 2006 results, i.e. that the cab-based screen was producing better training outcomes than the wide-screen. In addition to making fewer errors on both Trial 1 and Trial 2, the drivers using the cab-based screen also made larger improvements between the two trials as seen in Table 1.

However, the 2 (simulator) × 4 (decision point) × 2 (trial) mixed ANOVA performed did reveal a significant main effect for decision points (F = 17.29, df = 1.31, p = 0.000). The main effect of decision points indicates that almost all of the error rates were all significantly different from each other which means that the error rate for the Track Authority Work condition (24%) was significantly higher than the error rate for the Children on Track decision point (21%) which was significantly higher from the Failed Level Crossing decision point (10%) but Failed Level Crossing decision point error rate was not significantly different from the error rate in the Absolute Signal at Stop decision point (8%).

5.5. Presence

Despite the difference in skill levels achieved using the cab-based screen versus the wide-screen; the drivers rated the two simulators as equal on the presence/immersion measures. Both simulators scored 69% on the PQ and 61% on the IPQ as seen in Table 3. Additionally, both simulators were rated equally by drivers across all of the four factors underlying the PQ. There was a greater range between the highest and lowest scores for both questionnaires received for the cab-based simulator. However, the greatest deviation of scores for individual factors occurred with involvement and adaptation for the cab-based which was also the case for the wide-screen.

6. Discussion

In terms of the overall training program, the findings revealed performance did improve significantly when drivers were given the opportunity to repeat their simulator experience on the same day. This suggests that providing a repeat training session in the simulator may have had a cumulative effect on training effectiveness. Practice should result in greater confidence and improved performance. However, the results also indicate that, although performance improved across time within a day, retention over longer periods was not so good. Despite a year’s additional driving experience in the interim, drivers returning in 2007 performed significantly worse than during their training in 2006. Clearly some of the skills tested in the simulator involve circumstances which the drivers would rarely if ever experience in the real world, and this lack of repeated exposure may explain...
why some of the skills degraded over time but does not explain why they degraded to a point below the 2006 level.

Drivers were given a standard 5 min warm-up before each simulator exposure so it is unlikely that the difference in skill levels attained were a result of differing opportunities to access shorter or longer warm-up times. However, there may be a number of explanations for this apparently counter-intuitive result. It may be that classroom revision time undertaken prior to the simulation test differed dramatically from one year to the next. This time was not examined as part of this study. If revision was more targeted one year to the exact problems the drivers would encounter during their simulation session this would have provided much better preparation time. Alternatively, perhaps the fact that drivers were able to self-select the simulator they felt most comfortable with and the driving scene they undertook was unchanged from the previous year, this resulted in over-confidence which translated into risk-taking in the simulator and consequently more errors and mistakes. In their second attempt in 2007, the drivers demonstrated they did have high skill levels. This begs the question as to whether the improvement was a product of the practice in the first trial or whether in fact the drivers were more focused as a result of having made far too many errors on their first attempt.

The role of spaced repetitions in learning theory may also provide a future research direction to explore in an attempt to explain why skills decayed. There is very little research on how procedural learning takes place in simulated environments. However, there has been a great deal of research on how different spacing of repetitions in time affects the strength of memory and how the resulting findings could be applied in the practice of effective learning (Wozniak 1995). Memory can be strengthened through appropriate spacing time intervals; however, programs must be careful the spacing of training repetitions does not leave so much time between sessions that the learned memory trace becomes completely inaccessible. This theory would suggest it is highly probable that altering the training program in a simulator may assist to address the skill decay problem. In future research, it would be interesting to investigate how repetition after varying time intervals either improve or degrade retention specifically when simulators are in use. In large companies when relocating staff to central simulators is expensive it is particularly important to determine the longest inter-repetition interval that avoids retrieval failures.

When comparing learning effects across simulator type, the cab-based simulator produced higher decision making performance scores than the larger wide-screen simulator. This was surprising as it is generally accepted that higher-end immersive interfaces are more successful in rendering scenes similar to real-world settings. It has been high-end immersive facilities that have demonstrated that simulation works. However, researchers who have been working to increase understanding of immersion and identify the components that produce them acknowledge there is still far to go (Bowman and McMahan 2007). So what might the reasons be for this apparently counterintuitive discrepancy? Is it simply the case that occasionally less is more? Possibly, a smaller screen and smaller simulator might help aid concentration and reduce distraction. On the other hand, the cab-based simulators are in one sense closer to a real train. Their narrow field of view provides a visible frame of reference, something lacking in the large curved screen simulator. It may be that this frame of reference contributes to the realism and fidelity of the simulation, albeit covertly (as these differences did not emerge in the presence questionnaires).

Equally so it could also be that the perceptual field scanning for the wide-screen simulator is larger than is ever done in a real train, and drivers have neither the time, under the pressures of training for degraded conditions, nor the experience to scan effectively a ‘larger than life’ presentation. The rail training provider had installed wide-screen simulators to enable teams of people to be trained in team communication scenarios. However, the aim of multi-use for the simulators may have resulted in a detrimental impact for the solo driver training. A high perceptual immersion achieved via the full visual field may have been achieved at the cost of validity where the actual visual experience does not represent the real-world setting.

Both types of simulators used in the study achieved relatively low presence scores indicating drivers did not experience the virtual world as highly engaging. One of the main issues that drivers had was that they ‘...had no sense of train momentum and no feel of going up or down grades’. Drivers reported being able to adapt quickly to the controls and interface but the discrepancy between how natural the simulator felt in comparison to a real train may have been significant enough to reduce the immersive impact of the experience for some drivers.

Beyond specific issues of simulator design, the results of this study have highlighted a potential problem with the use of presence questionnaires in predicting training outcomes. Despite the significant difference in number of errors made when training was undertaken in the cab-based simulator versus the wide-screen simulators, the drivers rated the two simulators as equal on the presence/immersion measures. The presence analysis indicates that in terms of
subjective feedback neither the cab-based nor the wide-screen simulators were experienced as significantly more or less immersive or engaging. The main difference between them was that the scores for the cab-based simulator were more variable and this system was rated slightly more highly in terms of the interface which it offers.

In terms of having a stress inoculation effect, in the absence of high levels of presence, it is unclear if skill improvements between training sessions on the same day were a product of increased confidence in coping with high stress events or whether it was merely the result of an increasing level of comfort in driving through that particular scenario. Retesting drivers using a new track route along which the same problematic events (decision points) are embedded but in a different sequence would reveal whether drivers do maintain skill levels when under stress but without the benefit of familiarity.

7. Future directions

In the future it would be interesting to see to what extent the visual reference frame is responsible for the discrepancy in performance between the two simulators. It would also be helpful to test alternative measures of immersion and presence to see if they correlate more accurately with the training outcomes described here. The low presence levels did not predict the good training outcomes achieved on the simulators. In addition to enhancing performance, presence is also purported to be a strong indicator of whether training will ultimately transfer to the real world. If the higher performance scores are being influenced by levels of familiarity with the scene it could be that the lower presence scores should be more closely studied as an indicator of training transference. To investigate the effectiveness of the simulator training in the real world one approach would be to compare the annual incident rates for the year before training and the year after training for those drivers who have undergone simulation training. Ideally, a control group of drivers who did not receive any simulator training would also be compared against those who have. Similarly, incident rates of drivers who trained on the flat cab-based versus the curved wide-screen simulators could be investigated to determine if simulator type has an impact in the real world.

There is an important role for interactive simulators in replicating degraded events and establishing them as a core component of rail training programs. However, any simulator of any level of complexity can only be as good as the educational program in which it is embedded. Many simulators are underutilised due to failure to integrate them into a curriculum that enhances their potential. Future research investigating the optimal repetition spacing for simulator training sessions would contribute greatly to knowledge on how long memory traces built from virtual experiences last before requiring refresher training to ensure they remain accessible to personnel relying on the training to maintain safe operations.

Acknowledgements

The financial support of the Australian Research Council is gratefully acknowledged.

References


