The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology

http://dms.sagepub.com/

A review of the effectiveness of game-based training for dismounted soldiers

Susannah J. Whitney, Philip Temby and Ashley Stephens

The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology published online 25 January 2013
DOI: 10.1177/1548512912472773

Original article

JDMS

Journal of Defense Modeling and Simulation: Applications, Methodology, Technology
1–10

DOI: 10.1177/1548512912472773
dms.sagepub.com
SAGE

A review of the effectiveness of game-based training for dismounted soldiers

Susannah J. Whitney, Philip Temby and Ashley Stephens

Abstract

Computer games are increasingly being used by armed forces to supplement conventional training methods. However, despite considerable anecdotal claims about their training effectiveness, empirical evidence is lacking. This paper critically reviews major studies conducted in the past decade that have examined game-based training with dismounted soldiers. The findings indicate that these studies are characterized by methodological limitations and that the evidence regarding the effectiveness of game-based training for this military population is not compelling. Furthermore, due to methodological limitations with the studies, the possibility of negative training effects cannot be discounted. The paper concludes with implications for the scientific and military communities, as well as recommendations for the conduct of future studies in this area.

Keywords

Simulation, military training, evaluation, computer games

1. Introduction

This work is a critical review of the empirical evidence of the effectiveness of computer games in training dismounted soldiers. This contribution is important for two reasons. Firstly, to the best of our knowledge, this is the first such critical review to be undertaken; hence this is an original contribution. Secondly, the results of the review demonstrate that, in contrast to some of the anecdotal claims that have been made, the available empirical evidence on the training effectiveness of computer games for dismounted soldiers is very limited and comes from a small number of studies, the majority of which appear to have significant methodological limitations.

The most closely related work to this paper is Hays' review of the effectiveness of games for training.[1] However, this paper extends Hays' work through a more detailed review of the methodology and results of each study. In addition, this paper specifically addresses military training issues. This paper also provides a counter argument to Roman and Brown,[2] who suggest that the effectiveness of computer games for military training has been established. On the basis of this review, we conclude that the evidence for the effectiveness of computer games for training dismounted soldiers is weak at best.

This work benefits other researchers by highlighting the current state of evidence on the training effectiveness of computer games and by providing recommendations for future studies, to help avoid some of the methodological limitations affecting previous studies. The ultimate aim of research in this area is to provide robust evidence to military stakeholders to ensure informed decisions are made regarding the acquisition and use of computer games for training purposes.

Computer games are increasingly being used by armed forces to supplement conventional methods of military training, such as classroom instruction and field exercises. Furthermore, the military is investing heavily in the development of games for training purposes. For example, the United States (US) Army has committed to the investment

Defence Science and Technology Organisation, Edinburgh, South Australia, Australia

Corresponding author:
Susannah J. Whitney, DSTO, LOD Building 75, PO Box 1500, Edinburgh SA 5111, Australia.
Email: susannah.whitney@dsto.defence.gov.au

of $50 million on computer games to prepare soldiers for combat as part of a 'games for training' program.[3]

There are numerous claims in the literature about the potential benefits of computer games as military training tools.[2,4,5] These reported benefits include low purchase costs, reduced training times, reduction in live training costs, the ability to customize training scenarios, and the ability to practice skills that might otherwise be too dangerous or expensive to train on a regular basis by conventional means. Furthermore, the use of computer games as training tools has been widely advocated. For example, Roman and Brown (p10) state that they are an ''effective means to meet a wide variety of tactical training requirements''.[2] Based on claims such as these, one might assume that the training effectiveness of computer games has been well established. However, this assumption is not well supported at this time.

Despite the reported benefits, the majority of claims about the training effectiveness of computer games appear to be based on anecdotal evidence. Training effectiveness is defined in this paper as the extent to which skills learned during game-based training transfer to mission-specific activities and field exercises; this definition is consistent with that adopted by other authors in this area.[6,7] As highlighted by other researchers,[8–10] anecdotal evidence is insufficient to demonstrate training effectiveness, and empirical studies are required. However, few empirical studies on game-based training with military populations have been published. At present, armed forces seem to be adopting computer games for training purposes at a rate faster than the scientific community is able to publish sound evidence regarding their effectiveness. Without such evidence, there is a risk that military forces may be using these games in inappropriate ways or may be failing to realize the full range of benefits and return on investment. In the worst case, it is possible that game-based training has a negative impact on trainee performance and overall training outcomes, negating any cost-benefit associated with their use.

While more general reviews of game-based learning have been published,[1,11,12] there have been no critical reviews of game-based training with dismounted forces. The purpose of this paper is to examine the evidence regarding the effectiveness of game-based training for dismounted soldiers by critically reviewing major studies conducted in this area over the past decade. (The scope of this paper has been limited to dismounted soldiers as the authors have primarily been conducting research with this military population and are therefore more familiar with the published literature on game-based training with this group.) In doing so, the paper contributes to the body of knowledge on game-based training effectiveness in military populations, and identifies areas for future research.

2. Overview and organization

This paper critically reviews five major studies that have been conducted in the past decade.[13–18] Each study examined game-based training with dismounted soldiers at the individual and team level. This paper assesses the strength of the evidence that these studies provide in support of the effectiveness of game-based training for this population of military personnel.

At present, there is a variety of terminology in the published literature to describe 'computer games' and their use for training and instruction purposes. For example, the literature refers to computer games as 'games',[14,19] 'computer-based games',[17] 'virtual environments',[15] 'simulators',[16] 'virtual reality',[16] and 'serious games'.[2,20] It is beyond the scope of this paper to address the lack of consistency in terminology across studies and we simply highlight this point here for the reader. Despite this lack of consistency in terminology, it was still possible to identify studies where computer games were used for military training, which is the focus of this review. The use of computer games for military training is consistent with the term 'game-based training'. [1,10,19]

The scope of this paper is limited to evaluations of desktop computer games where outcomes from game-based and conventional training have been compared (in accordance with our definition of training effectiveness). Consequently, this review does not cover studies which only examined the use of games for experimentation, studies examining game-based training for mounted soldiers,[10,21] studies where performance following game-based training was not examined in the live, or field environment,[4,22–25] or studies where computer-based instruction was compared with traditional methods of knowledge acquisition in military populations.[7,26,27] The reader is therefore referred to these and other papers for broader coverage of game-based training.[1,11,28]

The paper is organized as follows. In the first section, the method by which the studies were located is briefly outlined. In the second section, the five studies are individually reviewed, with specific reference to their methodologies and key findings. In the third section, we identify several factors common to many of the studies, and discuss how these factors affect evaluations of game-based training. In the next section, conclusions are made about the effectiveness of game-based training for dismounted military forces, along with implications for the scientific and military communities. Finally, recommendations for future research are outlined. In this review, it is argued that the studies provide only weak evidence regarding the effectiveness of game-based training with dismounted military forces, and that more rigorous studies are required. Overall, it is concluded that, based on the available data, the best that can be said about game-based training for dismounted soldiers is that there is no compelling evidence for its effectiveness, and the possibility of negative training effects cannot be discounted.

3. Method

A literature search was conducted using Google, Google Scholar, and PsychINFO search engines. Searches were also conducted through unclassified databases of the Australian Defence Science and Technology Organisation, the UK Defence Science and Technology Laboratory, the US Army Research Institute for the Behavioral and Social Sciences, and the Defense Technical Information Center. The following search terms were used: computer games, COTS, dismounted forces, infantry, military training, training effectiveness, and virtual environment. When a relevant paper was found, a further Google Scholar search was conducted to locate any papers citing it. Some papers were also obtained directly from authors via email requests.

The following exclusion criteria were then applied to the papers. Papers were excluded if (a) the full paper could not be obtained, (b) the technology or training device examined was not a desktop computer game (for instance virtual reality or head-mounted display systems,[29] (c) the paper described only theoretical or technical aspects of using computer games for training,[20] (d) the paper did not examine transfer of training to the live environment,[23] or (e) participants were not military personnel.[30]

This search strategy located a total of six papers, covering five separate studies, which are reviewed individually in the following sections. Due to the small number of papers meeting the above criteria, and the lack of detailed results in some of the studies, it was not possible to conduct a meta-analysis. Consequently, the review was limited to examining the methodology, results, and conclusions from each study. It may seem surprising that so few studies satisfying these criteria were found. While other papers and presentations on game-based training were found which made claims of effectiveness, they included no quantitative data in support of these claims. Hence, the findings could not be considered distinguishable from anecdotal evidence. It is also of note that only one of the studies in this review, that by Proctor and Woodman,[18] was published in a peer-reviewed journal; the remainder were published as postgraduate study theses or military research reports.

4. Review of game-based training studies

4.1. Pennell (2003)

Pennell compared the effectiveness of game-based and conventional methods of training urban operations maneuvers,[14] including building clearance, patrolling, casualty evacuation (CASEVAC), and explosive entry procedures. Game-based training was conducted using a modified version of the computer game *Half-Life*, known as Dismounted Infantry Virtual Environment (DIVE). The aim of the study was to investigate the utility of DIVE as a training tool at the small team level.

The terrain used in DIVE was a virtual replication of the terrain used for live training and assessment. Participants were 16 UK Army personnel with infantry and engineering backgrounds; most had less than 4 years of military experience. On the first day of the study, all participants completed 30 minutes of familiarization training on DIVE, followed by a 45 minute practical lesson on building entry procedures. The 16 participants were then allocated into four teams, each with comparable levels of military experience. All participants then undertook a 45 minute practical lesson on room clearance, followed by a baseline test of their building clearance skills in their four teams in the live environment. Following this, the four teams undertook different training progressions over two days on a range of tasks under the supervision and assessment of Subject Matter Experts (SMEs). Two teams (Teams 1 and 2) conducted training in patrol and CASEVAC tasks in DIVE, while the other two teams (Teams 3 and 4) conducted comparable training on these tasks in the live environment. The next day, Teams 1 and 2 conducted the same tasks in the live environment, while Teams 3 and 4 conducted these tasks in DIVE to assess the benefits of each training method. Following this, Teams 1 and 2 conducted 90 minutes of explosive entry training in DIVE as a section then completed a test on these skills in the live environment. Teams 3 and 4 undertook the same training as a section but in reverse order (i.e. live training first then DIVE test second). On the final day of the study, the teams completed a section-level task in the live environment in a different location to that of previous days.

During the study, the performance of both groups was measured in three ways. The primary measure of performance was the time associated with key events during each building clearance task (e.g. time of first enemy contact, timing of casualties). Secondly, after each training serial, participants were asked to rate their own performance and their perceptions of game-based training on structured questionnaires. Thirdly, scoring criteria were used by SMEs to assess participants' performance during the game-based and live training.

Although the small sample size ($n = 4$) precluded statistical analysis of the results, several key findings were reported. Firstly, it was noted that there were no significant performance differences between teams following different DIVE and live training progressions, and mission timings and outcomes were similar in the virtual and live environment. Secondly, as training progressed, there was a tendency for participants to give higher ratings to the realism of the activity and to their ability to demonstrate skills in the game environment. Finally, it was noted that the SMEs believed that there was positive transfer of skills, such as the use of suppressive fire and grenades, from the game to the field. Unfortunately, there was no objective data to support the subjective assessments of the game's effectiveness. Notwithstanding this, Pennell (p5) concluded there were ''indications of positive benefit (or at least, no disbenefit [*sic*]) of using DIVE to train at section level''.[14] Overall, the study findings suggest that the game may have been beneficial for training urban operations skills; however, this suggestion is based largely on subjective data. At best, it appears that the game allowed the rehearsal of building clearance procedures and that certain skills were observed to transfer from the game to the live environment.

4.2. Nolan and Jones (2005)

Nolan and Jones examined the effectiveness of the game *Delta Force: Black Hawk Down – Team Sabre* to train infantry squads in reaction to contact drills.[13] The study hypothesis was that ''multiplayer commercial 'off the shelf' first person shooter games can be effective for use by infantry squads as a low resource, high impact small unit training tool'' (p1). Participants in their study were 41 US Army personnel enrolled in a junior officer training course. The sample was divided into experimental and control groups, with each group containing two squads. The researchers measured each participant's prior military experience, frequency of computer game playing, and self-reported marksmanship skill levels. As these variables did not differ significantly between the control and experimental groups, the researchers concluded that any performance differences between the groups were more likely to have resulted from training effects rather than individual differences associated with these variables.

The control group (n = 21) received no training beyond that provided in the officer training course. The experimental group (n = 20) was divided into two squads, which received 8 hours of training on the game, including 6–7 structured missions against a simulated enemy. After each mission, participants completed a questionnaire on which they rated their performance in the game environment, including their ability to communicate with team members, maintain their position in the formation, and engage the enemy. At the conclusion of game-based training, participants in the experimental group completed an additional questionnaire which primarily assessed their perception of the game's training effectiveness.

All participants then completed a field exercise, in which their reaction to contact skills were assessed by an SME. At the end of the study, all participants completed a questionnaire on which they rated their own performance and the extent to which they felt the training had prepared them for the field exercise. There were three main findings from the study. Firstly, the questionnaire data showed that the experimental group's reported level of competence increased across missions; however, this finding should be interpreted with caution given that it is based on participant self-report data. Secondly, the majority of participants rated the game as providing effective training, although some aspects of the game's visual and audio cues were rated poorly. Again, this finding is based on self-report data and is consistent with the anecdotal claims made about the effectiveness of game-based training. Thirdly, the SME rated the experimental group better than the control group in their movement, awareness, and planning in the field exercises, although it was not reported if these differences were statistically significant. Even if the differences were significant, there are still methodological constraints that could have affected this finding, which we discuss later.

Overall, based on the questionnaire data and the SME assessment, the researchers concluded that ''using COTS software has the *potential* to be an effective low cost and accessible training tool for training infantry squad collective tasks'' (p86, emphasis added).[13] The evidence on which the conclusion of training potential is based is weak. As noted by Nolan and Jones caution needs to be used in interpreting the self-report data due to the tendency for participants to overrate their performance.[13] In addition, methodological weaknesses may have affected the reliability of the field training assessment. Firstly, the SME in the study was not blind to the identity of the participants in the control and experimental groups, as the SME was involved in the game-based training. This means that the possibility of assessment bias cannot be discounted and that the SME's assessment of each group's performance might simply reflect this. Secondly, the lack of a baseline performance measurement precluded an assessment of the extent to which learning took place as a result of the game-based training. Thirdly, it appears that the experimental group received more training time overall than the control group; hence the performance difference (significance aside) may simply reflect more time spent improving their knowledge of the military task being trained (i.e. reaction to contact). Each of these reasons could plausibly account for the SME's assessment of the two groups' performance. Hence, we believe the methodological shortcomings in this study preclude any strong conclusions being made regarding the game's effectiveness as a training tool, although the consideration of participant experience levels is a noteworthy methodological strength. Overall, however, the study provides little evidence regarding the effectiveness of using games to train dismounted soldiers in collective skills.

4.3. Wiederhold (2005)

Wiederhold examined the effectiveness of a computer game for training a range of tactical skills to military personnel in the US Navy, US Marine Corps, and US Coast Guard taking part in a training exercise.[15] The study's objectives were to ''examine the effectiveness of virtual reality (VR) training simulators regarding their ability to teach tactical and trauma care skills, to practice stress management techniques, and to improve performance during real-life combat situations'' (p1). (While this study purports to examine the effectiveness of VR technology, the technology used in the study was a laptop based, first person simulation, and so met our criteria for inclusion in this review.)

A total of 970 participants from five different units participated in the study. Three of the five units conducted training in urban tactics, techniques and procedures. The other two units conducted training in boarding, searching, and seizing a ship. Due to space limitations and the scope of this paper, only the results from one US Marine Corps unit are reviewed. However, the methodology and findings were similar across units.

In this unit, 210 participants were allocated into either an experimental (n = 90) or control group (n = 120). The experimental group received game-based training consisting of 15 minutes of navigation through a virtual shoothouse and 15 minutes of navigation through a virtual village. The virtual shoothouse and village were accurate replicas of the live training facilities. The control group received no such training. Both groups then conducted live training exercises in a real shoothouse and village. Live training included a range of tasks, including entering, clearing, and securing the shoothouse, locating, recovering, and evacuating casualties, and exfiltration from the village.

The performance of each group was assessed using a number of subjective and objective measures, including SME observation and the time taken to clear rooms. The total duration of the training program was 11 days. Although the time spent training on each particular task was not specified, it can be deduced that the time spent training on the simulator constituted only a small fraction of the total training time.

Wiederhold reports that the experimental group outperformed the control group on all subjective and objective performance measures.[15] For example, he notes that SMEs judged that the experimental group took better cover when approaching the shoothouse, and demonstrated quicker and more coordinated movement. The experimental group was 2 seconds faster than the control group in room clearance, and 2 minutes faster in securing a building. Wiederhold reports that these differences were statistically significant, but does not report the statistical tests used. Other than for timing data, there is no indication as to whether any of the differences based on objective measures are statistically significant, and the results are not reported in sufficient detail to allow such an analysis. Based on the findings, the author concluded that ''the laptop simulator has proven to be an effective and efficient method of training. Skills obtained and polished in the trainer not only enhance mission specific behaviors but readily generalize to many other skills and behaviors'' (p33).

Wiederhold's conclusions seem ambitious,[15] given the methodological weaknesses of the study. Firstly, the findings are likely to be confounded by two factors, firstly the fact that the virtual environment was a replica of the live training environment (i.e. the real village), and secondly that the control group received no additional training. The use of a virtual environment that replicates the live environment is likely to have provided knowledge of specific features of the environment, such as which way a door opened, or the location of hidden rooms. This suggests that the virtual environment provided some terrain familiarization. However, as the control group received no additional training, it is impossible to conclude with any certainty if the training provided through the virtual environment was superior to other forms of training, or if training in a virtual environment that was not a replica of the live environment would have yielded any training benefit. We do not think that there is sufficient evidence to support Wiederhold's claims on training effectiveness, in particular, the claim regarding generalizability of skills.

Another concern is that there is no indication if the SMEs who assessed live performance were blind to the identity of the participants in the experimental and control groups. Moreover, there is incongruity between the skills trained in the virtual environment, and the skills tested in the live environment. For instance, participants were assessed in the live environment on their ability to search unknown personnel. There is no indication that this was trained in the virtual environment; given the training sessions lasted only 15 minutes, it is unlikely this – or any other skills – would have been learned in this time. The lack of skill improvement may also be due to the considerable military experience of many of the participants. As there is no indication how participants were allocated into the groups, and there was no baseline assessment prior to training, the possibility that performance differences between the two groups was due to *a priori* differences in experience cannot be discounted. Overall, the strongest conclusion that can be drawn from the study is that the use of the simulator was better than no training when this training involved conducting equivalent tasks to those subsequently assessed in the live environment.

4.4. Kneuper (2006)

Kneuper examined the effectiveness of the game *Delta Force: Black Hawk Down* to train a range of military tasks,[17] including ambush, reconnaissance, and small unit attack. The participants in the study were 80 US Army officer cadets who were completing a Reserve Officers' Training Corps program. Participants were assigned using stratified random sampling on the basis of gender, ethnicity, prior computer game experience, and academic ability to one of six groups corresponding to the amount of game-based training received (i.e. 0%, 15%, 30%, 45%, 60%, and 75%). All participants received a total of 20 hours of training. Participants who undertook game-based training (five out of the six groups) also received three hours of training on the game to become familiar with the controls and basic movements. Live and game-based training sessions were similar in structure; cadets received orders, planned, and executed their mission, and were then given feedback on their performance. A total of 50 participants, from the 80 who started the program, completed the training.

Following training, the 50 remaining participants were assessed by SMEs on 12 leadership dimensions over five days during a field exercise. The 12 leadership dimensions were: Mental, Physical, Emotional, Conceptual, Interpersonal, Technical, Tactical, Communication, Decision-Making, Planning, Executing, and Assessing. For each dimension, participants were assessed as either 1 = 'Needs Improvement', 2 = 'Satisfactory', or 3 = 'Excellent'. Participants were assessed several times during the field exercise and the mean scores for the six different groups were then calculated. The results indicated that there were no significant differences in the mean scores on any of the dimensions for the six groups. Kneuper suggested that the lack of significant differences between the groups was evidence that replacing live training with game-based training did not have a negative impact on the performance of the cadets.[17] In addition, Kneuper suggested that since the group that received 45% live and 55% virtual training had the highest mean score on 8 out of the 12 dimensions, this ratio represented the optimal mix of live and game-based training.

Kneuper's study is the only study to our knowledge to have explored the effect of varying amounts of game-based and live training on performance outcomes in a military context.[17] While the use of stratified random sampling is a methodological strength, the findings from his study are otherwise limited by significant shortcomings. First, the three-point rating scale used for each of the 12 dimensions was not particularly sensitive to detecting performance differences between participants. Second, as Kneuper reports, the assessors were biased towards giving 'Satisfactory' ratings, further reducing the likelihood of detecting performance differences. Third, there appears to be some incongruity between the skill sets being trained and the skill sets being assessed. As noted by Kneuper,[17] the focus of training was on leadership skills, including planning and decision-making. However, during the field exercise, participants were assessed on a broad range of attributes, including physical fitness and military bearing, self-control, and interpersonal skills. The incongruity between the skills being trained and assessed raises some doubt over the validity of the measures used in the study. Unfortunately, no psychometric data on the measures was provided by the researcher, so no further comments can be made about this. Fourth, there were no baseline measures collected from any of the participants prior to receiving training, so it is not known how individual and group performance changed as a result of the training.

Finally, a major concern in the Kneuper study is the stated conclusion that an optimum mix of live and game-based training can be inferred from the findings despite the non-significant results obtained.[17] This conclusion is of particular concern given that the researcher acknowledges that ''the data also seems to show that the evaluation schema is unable to really tell what is happening to the cadet's performance'' (p93). Overall, the strongest conclusion that can be drawn from Kneuper's study is that there is no evidence that replacing live training with game-based training had a negative impact on performance. However, this conclusion should still be treated with caution given the study's methodological limitations.

4.5. Woodman (2006)

In his study, Woodman examined the potential of the game *Close Combat: First to Fight* to train room clearance skills.[16,18] The null hypothesis tested was that ''traditional field training is equivalent to virtual training combined with field training'' (p34).[16] The participants were 32 US Marine Corps personnel enrolled in a close quarters battle training course. Participants in the control group (n = 16) received 3 hours of supervised study associated with the course, and 3 hours of mentored 'walk-throughs' of the test location. Participants in the experimental group (n = 16) received 2 hours of supervised study, 2 hours of walk-throughs, and 2 hours of structured training sessions using the game. The experimental group also received 45 minutes of exposure to the game to become familiar with the game features and controls. During the 2 hours of structured training, the participants completed a series of missions consisting of briefing, planning, execution, and after action review (AAR) phases.

Following training, all participants were assessed on their room clearance skills in an actual building. Each participant was assigned the role of team leader and three SMEs played the remaining members of the team. SMEs were blind to the type of training each participant had received. Participants' performance was evaluated by the SMEs using a checklist of training objectives. Participants in the experimental group also completed a questionnaire on the effectiveness of the game for training planning skills, situation awareness, team communication, tactical awareness, movement techniques, weapons employment, and rules of engagement.

The results of the SME assessment showed that there was no statistically significant difference between the performance levels of the two groups. Interestingly, this result meant that the null hypothesis (i.e. that traditional field training is equivalent to virtual training combined with field training) could not be rejected. This suggests that the two methods of training may have been equally effective; however, the researcher does not acknowledge this possibility. Instead, Woodman identified several factors that may have contributed to the null result.[16] Firstly, the experimental group only received two hours of structured training with the game, which limited the amount of time that participants had to rehearse tactical procedures. Secondly, the performance measures used in the study were too blunt, with certain procedural errors resulting in automatic fails or point deductions, and not sensitive to detecting differences between groups. Thirdly, the limited field of view within the game environment made it difficult for the participants to rehearse movements and maintain situation awareness.

In spite of these methodological weaknesses, the researchers concluded ''substitution of a portion of field training with game-based training did not harm the performance of Marines either objectively or subjectively when compared to the performance of Marines receiving regular field training alone'' (p61).[18] On the surface, this conclusion appears reasonable. However, in light of the study's methodological limitations, evidence regarding the game's effectiveness is weak, relative to the normal training method. The possibility that the 2 hours of study and 2 hours walk-throughs the experimental group received was of similar benefit to the 3 hours of study and 3 hours of walkthroughs the control group received cannot be discounted. If so, this would mean that the two hours of game-based training received by the experimental group had no training benefit at all.

Moreover, because a pre and post-test design was not employed, it is not possible to determine if the different types of training each group received resulted in different changes in performance. It is possible that the two hours of walk-throughs the experimental group received was of similar benefit to the three hours the control group received – suggesting that the 2 hours of game-based training received by the experimental group had no training benefit at all. Furthermore, the question could be asked: why go to the trouble of using a game for the sake of 2 hours when you could train in the actual facilities in that time at no additional cost? Overall, we conclude that the findings from Woodman's study provide little compelling evidence regarding the effectiveness of the game for training room clearance skills.[16]

Despite the above issues, there are three strengths in Woodman's study that are worthy of mention,[16] although overall they are unlikely to compensate for the previously mentioned weaknesses. Firstly, the study used blind methods of assessment; this was the only study in those reviewed that used this approach. Secondly, standardized assessment criteria were used and participants were rated on a checklist containing 27 items. Using multiple items increases the chances of detecting performance differences, even when an aggregate score is used to measure overall performance. Thirdly, the same SMEs were used to assess all of the participants; such an approach increases the likelihood that ratings will be consistent and reliable.

5. Discussion

In reviewing the studies, a number of common issues and themes were apparent; these are now briefly discussed.

Firstly, in most studies, there was no significant difference in the training outcomes for participants who received a combination of conventional and game-based training when compared with participants that received conventional training only. No study found any evidence that game-based training produced superior outcomes than conventional training alone; rather there was no clear evidence of negative training effects.

Secondly, it was often difficult to determine exactly how the study was conducted because insufficient details were provided in the studies or the methodology was not well described. In particular, studies frequently failed to describe in detail the specific performance measures employed and how training was conducted. In addition, in some instances, the researchers failed to specify research hypotheses making it virtually impossible to interpret the outcomes and draw meaningful conclusions.[14,15] Conversely, when hypotheses were specified, there was a tendency for researchers to not interpret their findings in the context of these hypotheses.[16] Such lack of detail limits the ability of other researchers to replicate studies and assess the validity of outcomes from published studies.

Where a null result was obtained, this appeared to be interpreted as a negative outcome, and attributed to methodological weaknesses.[16] Overall, there appeared to be a tendency for researchers to eschew making inferences about what non-significant results actually meant in the context of their research; the implicit conclusion appeared to be that game-based plus conventional training is equally as effective as conventional training alone. However, as noted by Boldovici et al.,[31] to assume that a null result implies equal effectiveness is flawed reasoning because other factors, such as statistical power and effect sizes, can contribute to the finding of a null result. As none of the studies reported effect sizes for any of their variables, no statements can be made about the magnitude of the treatment effects.

Thirdly, another feature of these studies was that the measures of effectiveness were primarily based on subjective data from participants and SMEs. In all studies, no baseline data was collected and, with one exception, blind evaluation methods were not employed. As a consequence, it is not possible to make statements about changes in participants' performance from pre-training levels, nor is it possible to rule out assessment bias. (The majority of studies reviewed were conducted in conjunction with military training events, which can pose considerable challenges for researchers, as noted by Salas et al.[32] Consequently, the issues outlined in this paper may arise, in part, as a result of this factor.)

Fourthly, another common finding was that, in many studies, the game was incapable of providing sufficient fidelity to support certain training objectives, such as the ability to make precise movements, check for casualties, and maintain situation awareness. Consequently, the extent to which these skills could be adequately trained appeared limited.

Another common finding was that small amounts of time were allocated to game-based training. While researchers are often constrained by time and resources when conducting studies as part of military training programs, limited amounts of training reduce the chances of observing learning effects, and compromise rigorous evaluations of training effectiveness.

Small to moderate sample sizes and non-homogeneous groups were another consistent feature of the studies reviewed. Having too few participants reduces the likelihood of detecting genuine differences between groups (i.e. statistical power) and non-homogeneous samples prevent researchers from being able to generalize the findings beyond their studies. Non-homogenous samples also reduce the ability to draw inferences about the relationship between the type of training method and training outcomes because the outcomes may be confounded by individual (within-group) differences, such as length of military service and prior experience with computer games. While these variables can be controlled to some extent in post-hoc analyses, it is preferable to use participants with similar backgrounds and military experience. In the latter case, it is more likely that learning or training effects will be observed with less experienced samples. Some of the studies included participants with significant military experience, which may have had a negative impact on the evaluation outcomes. When studies used less experienced participants, the amount of training time was likely to be insufficient to support learning outcomes.[15]

Finally, there was a tendency for researchers to draw invalid conclusions from their results and to overstate the impact of their results.[17] To do so increases the risk that the claims will be perpetuated by practitioners and other researchers, which is potentially detrimental to the military and scientific communities. Overall, as suggested by De Freitas,[11] there is a need for more rigorous baseline studies to quantify how much and in which ways games are being used to support learning.

6. Conclusion

This paper has reviewed major studies on game-based military training involving dismounted soldiers that have been conducted in the past decade. The purpose of this review was to examine the evidence regarding the effectiveness of game-based training with this population of military personnel. While game-based training can potentially reduce training costs and improve the efficiency of military training, the empirical evidence in support of game-based training is weak, and hence its effectiveness for training dismounted soldiers cannot be concluded with any certainty at this time. In the majority of studies, a common finding was that the performance of military personnel receiving traditional training was indistinguishable from the performance of military personnel receiving a combination of conventional training and game-based training. While this finding might be taken as evidence of equal effectiveness (i.e. that a combination of game-based and conventional training produces equivalent outcomes to conventional training), it is equally plausible that this finding is due to the effects of conventional training alone, especially given the small amounts of time allocated to game-based training in most studies, and the varying experience levels of participants in some studies. In the absence of baseline data for game-based training (versus conventional training), we can only speculate as to which of these arguments is most likely.

Overall, based on the findings from this review, there is little compelling evidence at this time for the effectiveness of game-based training for dismounted forces. Furthermore, in light of limitations associated with a number of the studies, the possibility that negative training is occurring cannot be discounted. Given the considerable investment that is being made into game-based training, there is a need to ensure that these games deliver effective training. Otherwise, there is a risk that game-based training for dismounted forces will provide limited (or no) return on investment. Where game-based training is not effective, traditional methods of instruction (e.g., classroom lessons, walkthroughs) should not be overlooked as they may be equally or more effective than game-based training as well as being less expensive.

For the scientific and military communities, the question remains: to what extent can conventional training be *replaced* by game-based training? This question is certainly an area worthy of future research, along with those areas outlined in the following section. Another implication for the military community is that for future research to provide answers to these questions, then significant resources will need to be made available to researchers to undertake more evaluation studies. Overall, we believe that more studies, based on rigorous experimental methods, are needed to properly determine the effectiveness of game-based training and the appropriate mix of live and virtual training.

7. Recommendations for future research

Given the issues identified in this review, we suggest a number of recommendations for future research. These recommendations are based on established scientific principles and practices so they will come as no surprise to many. However, they are documented here as a reference for future researchers. (The article by Boldovici et al. also contains a list of useful recommendations for researchers to consider when conducting training evaluation research.[31]) First, it is recommended that researchers clearly articulate their *research objectives, questions and hypotheses* and relate their findings back to these objectives. This information greatly assists other researchers when judging the success of a study in relation to how the study was conducted, the type and amount of data collected, and the reliability of the research outcomes. Needless to say, it also provides the rationale for conducting the research in the first place. More generally, it is recommended that researchers include sufficient detail in their reports to allow replication of methodologies and to assist researchers in interpreting the outcomes. In particular, researchers should document effect size data and confidence intervals, to assist future studies, such as meta-analyses.

Second, it is recommended that researchers identify the type of training device being used and define the *types of skills* that are being examined, in order to assess the extent to which the device can support the training of those skills. It is also recommended that researchers conduct studies over longer periods to examine the retention of these skills over time, including those required to play computer games.

Third, it is suggested that researchers give more consideration to the *statistical power* of their studies by ensuring that adequate sample sizes are employed; while this issue can be problematic when studies are conducted in conjunction with military training events, one solution is to conduct multiple studies with different cohorts.

Fourth, it is suggested that researchers give consideration to their *sampling methods* in the planning stages of their studies. Typically, in training evaluation studies, researchers have little influence over the selection of participants and how they are assigned to different treatment conditions. Unfortunately, such sampling methods can reduce the validity of the claims about training outcomes, and therefore it is important that future research is conducted with more representative samples (e.g. similar levels of computer game and military experience). Therefore, future research is likely to require increased levels of support from the military to ensure appropriate samples are used.

Fifth, researchers should ensure that their *performance measures* are both reliable and valid, and sensitive enough to detecting changes in performance. It is also recommended that researchers collect baseline data on the skills under investigation prior to training interventions; this will allow more definitive conclusions to be drawn about the relative contribution of game-based training to overall training outcomes. It is also recommended that future studies employ blind assessment methods when evaluating performance to reduce the potential for assessor bias.

Sixth, it is recommended that the *amount of time* allocated to game-based training is given more consideration in future research; ensuring adequate amounts of training will reduce the likelihood that training outcomes are confounded by limited opportunities for learning. It is also recommended that better consideration is given to the treatment condition(s) assigned to control groups, and that, where appropriate, equal amounts of training time are allocated to all groups.

Finally, it is recommended that future studies be subjected to *rigorous peer review* so that the claims made in published papers can withstand the scrutiny of the scientific community. By addressing each of the recommendations above, it is our hope that future research will uncover the true effectiveness of game-based training for military and other populations.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Hays RT. *The effectiveness of instructional games: a literature review and discussion*. Orlando, FL: Naval Air Warfare Center Training Systems Division, 2005.
2. Roman PA and Brown D. Games – just how serious are they? In: *Interservice/industry training, simulation, and education conference*. Orlando, FL: NTSA, 2008.
3. Robson S. Not playing around: army to invest $50m in combat training games, www.stripes.com/article.asp?section=104&article=59009 (2008, accessed 18 October 2011).
4. Morrison P and Barlow M. Child's play? Coercing a COTS game into a military experimentation tool. In: *SimTecT*

2004. Canberra, ACT: Simulation Industry Association of Australia.
5. Prensky M. *Digital game-based learning*. New York: McGraw Hill, 2001.
6. Beal SA. Using games for training dismounted light infantry leaders: Emergent questions and lessons learned. Research Report 1841, US Army Research Institute for the Behavioral and Social Sciences, Arlington, VA, 2005.
7. Ricci KE, Salas E and Cannon-Bowers JA. Do computer-based games facilitate knowledge acquisition and retention? *Mil Psychol* 1996; 8: 295–307.
8. Belanich J, Mullin LN and Dressel JD. Symposium on PC-based simulations and gaming for military training. Research Product 2005–01, US Army Research Institute for the Behavioral and Social Sciences, Arlington, VA, 2004.
9. Knerr BW. Current issues in the use of virtual simulations for dismounted soldier training. In *NATO human factors and medicine panel workshop on ''virtual media for military applications''*, 2006, pp.21.1–12. West Point, NY: NATO.
10. Ratwani K, Orvis K and Knerr B. An evaluation of game-based training effectiveness: context matters. In *I/ITSEC 2010*. Orlando, FL: NTSA.
11. De Freitas S. *Learning in immersive worlds: a review of game-based learning*. Bristol, UK: Joint Information Systems Committee, 2006.
12. Sitzmann T. A meta-analytic examination of the instructional effectiveness of simulation based computer games. *Personnel Psychol* 2011; 64: 489–528.
13. Nolan JM and Jones JM. *Games for training: leveraging commercial off the shelf multiplayer gaming software for infantry squad collective training*. Monterey, CA: Naval Postgraduate School, 2005.
14. Pennell RM. Dismounted infantry virtual environment training effectiveness trial (DIVE TET) 03. Report TR030689, QinetiQ, Farnborough, UK, 2003.
15. Wiederhold MD. *Physiological monitoring during simulation training and testing*. Durham, NC: US Army Research Office, 2005.
16. Woodman MD. *Cognitive training transfer using a personal computer-based game: a close quarters battle case study*. Orlando, FL: University of Central Florida, 2006.
17. Kneuper GM. *Substituting live training with virtual training by means of a Commercial Off The Shelf, First Person Shooter computer game and the effect on performance*. Orlando, FL: University of Central Florida, 2006.
18. Proctor MD and Woodman MD. Training ''shoot house'' tactics using a game. *J Defense Model Simul* 2007; 4: 55–63.
19. Topolski R, Leibrecht B, Cooley S, et al. *Impact of game-based training on classroom learning outcomes*. Arlington, VA: US Army Research Institute for the Behavioral and Social Sciences, 2010.
20. Morrison P, Parr S and Ward D. The evolution of 1st-person trainers: a case-study with VBS and HLA integration. In: *SimTecT 2007*. Canberra, ACT: Simulation Industry Association of Australia.
21. Brown B. *A training transfer study of simulation games*. Monterey, CA: Naval Postgraduate School, 2010.
22. Barlow M, Morrison P, Luck M, et al. Constructing the virtual section. In: *SimTecT 2004*. Canberra, ACT: Simulation Industry Association of Australia.
23. Beal SA and Christ RE. Training effectiveness evaluation of the Full Spectrum Command game. Technical Report 1140, US Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA, 2004.
24. Singer MJ and Knerr BW. Evaluation of a game-based simulation during distributed exercises. Research Report 1931, US Army Research Institute for the Behavioral and Social Sciences, Arlington, VA, 2010.
25. van der Hulst A, Muller T, Besselink S, et al. Bloody serious gaming – experiences with job oriented training. In: *I/ITSEC, 2008*. Orlando, FL: NTSA.
26. Parchman SW, Ellis JA, Christinaz D, et al. An evaluation of three computer-based instructional strategies in basic electricity and electronics training. *Mil Psychol* 2000; 12: 73–87.
27. Stone R, Caird-Daley A and Bessell K. *SubSafe:* A games-based training system for submarine safety and spatial awareness (Part 1). *Virtual Reality* 2009; 13: 3–12.
28. Caspian Learning. *Serious games in defence education*. Sunderland, UK: Caspian Learning, 2008.
29. Knerr BW and Lampton DR. *An assessment of the virtual-integrated MOUT training system (V-IMTS)*. Arlington, VA: US Army Research Institute for the Behavioral and Social Sciences, 2005.
30. BinSubiah A, Maddock S and Romano D. A practical example of the development of a serious game for police training. In: Ferdig RE (ed) *Handbook of research on effective electronic gaming in education*. Hershey, PA: Information Science Reference, 2008, pp.451–477.
31. Boldovici JA, Bessemer DW and Bolton A. *The elements of training evaluation*. Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences, 2002.
32. Salas E, Milham L and Bowers C. Training evaluation in the military: misconceptions, opportunities, and challenges. *Mil Psychol* 2003; 15: 3–16.

Author biographies

Susannah J. Whitney is a Cognitive Scientist at the Australian Defence Science and Technology Organisation (DSTO). She holds a Bachelor of Arts Degree (with First Class Honours) from the University of Newcastle, and a PhD in Psychology from the University of Queensland.

Philip Temby is a Senior Research Scientist at the DSTO. He holds a Bachelor of Science Degree (with First Class Honours) from The University of Adelaide, and a Master of Psychology Degree from University of South Australia. He is also a Registered Psychologist in South Australia.

Ashley Stephens is an Operations Analyst at the DSTO. He holds a PhD in Physical and Inorganic Chemistry from the University of Adelaide.