Australian Government
Department of Defence
Defence Science and
Technology Organisation

Evaluating the Effectiveness of Game-Based Training: A Controlled Study with Dismounted Infantry Teams

Susannah J. Whitney, Philip Temby and Ashley Stephens

Land Operations Division
Defence Science and Technology Organisation

DSTO-TR-2799

ABSTRACT

Computer games are increasingly being used by armed forces to supplement traditional methods of military training, despite a lack of empirical evidence on their training effectiveness. This report describes a study conducted by DSTO scientists examining the effectivenesss of a desktop computer game to train small teams of dismounted soldiers in infantry tactics, techniques, and procedures. One infantry section received traditional field-based instruction in section attack procedures, the other took part in game-based training using Virtual Battlespace 2. The performance of both sections was measured before, during, and after training. While the performance of the field-based training section improved significantly from pre-training to post-training, the game-based section showed no significant changes in performance. Overall the findings suggest that game-based training is not effective for training novice teams of infantry personnel in section attack procedures; this contrasts with several previous studies which found game-based training to be effective. The implications for using desktop computer games for individual and team training are discussed and recommendations for future research in this area are outlined.

RELEASE LIMITATION

Approved for public release

Published by

Land Operations Division
DSTO Defence Science and Technology Organisation
PO Box 1500
Edinburgh South Australia 5111 Australia

Telephone: (08) 7389 5555
Fax: (08) 7389 6567

AR-015-508
January 2013

APPROVED FOR PUBLIC RELEASE

Evaluating the Effectiveness of Game-Based Training: A Controlled Study with Dismounted Infantry Teams

Executive Summary

Computer games are increasingly being used by armed forces to supplement traditional methods of military training. While the potential benefits of these games are well documented; there is little objective evidence to support their perceived training benefits, especially in the area of collective training for dismounted soldiers.

This report describes an experiment examining the effectiveness of Virtual Battlespace 2 (VBS2) in training section attack tactics, techniques and procedures (TTPs). The study was conducted by Defence Science and Technology Organisation (DSTO) scientists under Task ARM 07/163 Training and Preparedness in response to a request from Training Command - Army to evaluate the efficacy of game-based technologies within the Australian Army, and to provide advice on how to best use these tools for training. The aims of the study were to: (a) compare training outcomes for game-based and field-based training; (b) evaluate the effectiveness of VBS2 for training novice infantry teams in section attack; (c) test a methodology for evaluating game-based training; and (d) provide advice to the military customer regarding the study findings and implications.

Two infantry sections took part in the study and were allocated to either game-based or field-based methods of training section attack procedures. Prior to training, a baseline assessment of each section's ability to conduct section attack in the field was undertaken. Following this, one section received eight hours of game-based training using VBS2; the other section received eight hours of field-based training in section attack procedures. All training was conducted by experienced Section Commanders. After training, both sections were again assessed on their ability to conduct a section attack in the field. All assessments of section attack performance were conducted by military Subject Matter Experts (SMEs).

The key findings from the study were as follows:

- Game-based training for section attack was not effective. The performance of the game-based section did not change significantly after training from their pre-training levels. There was no significant evidence that section attack skills learned during game-based training transferred to the field.

- Field-based training for section attack was effective. The performance of the field-based section improved significantly after training. This evidence suggests that the current method of training section attack to novice infantry soldiers is effective.

These outcomes are in contrast with previous studies with dismounted combatants, which have concluded that game-based training is effective when delivered in addition to or in combination with traditional training methods. However, these studies were unable to quantify the relative effectiveness of game-based training and traditional methods, as we have in the current study.

The most likely reasons for the current findings are: (1) the difference between the physical skills used in the VBS2 virtual environment and those used in the real world, and (2) limitations with the game's virtual environment such as the restricted field of view (which impacted on the ability of the section to maintain situation awareness) and difficulties in target indication, which resulted from poor audio cues. Possible solutions to these current shortcomings could be investigated by future research (e.g., improving the field of view, and developing some form of automatic target indication). It is possible, however, that such solutions may not be sufficient to improve the effectiveness of VBS2 for training this type of military task; the difference in the physical skills used in the virtual environment and those used in the real world (i.e. field environment) may still be the limiting factor.

The key implications of the findings for Defence are that (1) game-based training with VBS2 would not be considered a cost-effective alternative to the current method of training dismounted soldiers in section attack procedures, and (2) the results are likely to generalise to other collective skills/tasks undertaken by dismounted soldiers, including potentially the preparation of trained soldiers prior to deployment. While the current findings do not support the use of VBS2 for training section attack procedures, it is possible that such computer games might be effective for training other military tasks, such as mission rehearsal and terrain familiarisation. Consequently, it is suggested that alternative applications for such tools be investigated and evaluated as part of future research.

Recommendations for future research are as follows:

- Identify the military tasks that are most suitable for game-based training.
- Examine the optimal mix of game-based and traditional training.
- Conduct all future studies in this area with a similar level of scientific rigour to that employed in this study.

Authors

Susannah J. Whitney
Land Operations Division

Susannah Whitney holds a PhD in Psychology from the University of Queensland, and a Bachelor of Arts (Psychology) with First Class Honours from the University of Newcastle (NSW). She joined DSTO in February 2007 as a Cognitive Scientist. Prior to commencing at DSTO she worked as a lecturer in Psychology and Human Factors at the University of Queensland and Queensland University of Technology. Her research includes the training effectiveness of simulation and digital literacy.

Philip Temby
Land Operations Division

Philip Temby is a Research Scientist in DSTO's Land Operations Division. He holds a Bachelor of Science Degree and First Class Honours Degree in Psychology from the University of Adelaide. He also holds a Master of Psychology Degree from the University of South Australia and is a Registered Psychologist in Australia. He joined DSTO in 2002 and has conducted human factors studies in support to various Australian Army and Special Forces projects. His research interests include skill learning and development, team training, performance measurement, and personnel selection.

Ashley Stephens
Land Operations Division

Ashley Stephens obtained a PhD in Physical and Inorganic Chemistry from Adelaide University in 1995. He undertook postdoctoral studies at the University of British Columbia from 1995 to 1998, and at Flinders University from 1998 to 2000. He joined DSTO in 2001 and has conducted Operations Analysis to support a range of tasks including the AEF, Urban Operations, Special Operations Capability Support and Training and Preparedness.

This page is intentionally blank

Contents

GLOSSARY

1. INTRODUCTION 1
1.1 Background 1
1.2 Previous Research 2
1.3 The Current Study 6
1.3.1 Research Questions and Objectives 6
1.3.2 The Infantry Section 6
1.3.3 Section Attack 7
1.3.4 Virtual Battlespace 2 8

2. METHOD 9
2.1 Study Design and Ethical Approval 9
2.2 Participants and Assessors 9
2.3 Equipment 10
2.3.1 Computer Hardware and Software 10
2.3.2 Weapon Systems 10
2.4 Measures 11
2.4.1 Background Questionnaire 11
2.4.2 Section Attack Assessment Criteria 11
2.4.3 AAR Data 13
2.4.4 Exit Questionnaire 13
2.5 Procedure 14
2.5.1 Introductory Brief and Background Questionnaire 14
2.5.2 Allocation to Sections 14
2.5.3 Theory Lessons 15
2.5.4 Pre-Training Assessment 15
2.5.5 Section Attack Training 15
2.5.5.1 Game-Based Training 16
2.5.5.1.1 Game Familiarisation 16
2.5.5.1.2 Training Missions 17
2.5.5.2 Field-Based Training 18
2.5.6 Within-Training Assessment 18
2.5.7 Post-Training Assessment 19
2.5.8 Exit Questionnaire and Debrief Session 19
2.6 Data Analysis 19

3. RESULTS 20
3.1 Demographic Data 20
3.1.1 Length of Military Service 20
3.1.2 Computer Game Experience 20
3.1.3 Confidence in Conducting Section Attack Procedures 21
3.2 Pre-Training Data 21

3.2.1 Competency-Based Assessment 23
3.3 Training Data 24
3.3.1 Game-Based Training 24
3.3.2 AAR Data 25
3.3.3 Field-Based Training 26
3.4 Post-Training Data 26
3.4.1 Impact of Non-Blind Assessors 29
3.4.2 Competency-Based Assessment 30
3.4.3 Knowledge-Based Assessment 31
3.4.4 Exit Questionnaire Data 32

4. DISCUSSION 33
4.1 Comparison of Game-Based and Field-Based Training Outcomes 34
4.2 Evaluation of the Effectiveness of VBS2 for Training Section Attack 34
4.2.1 AAR Data 36
4.3 Methodology for Evaluating Game-Based Training 36
4.3.1 Comparison with Previous Studies 37
4.3.2 Comparison with TTCP GUIDEx 37
4.4 Implications of Findings 38
4.4.1 Cost-Benefit 38
4.4.2 Generalisability to Other Collective Skills 38
4.4.3 Collective versus Individual Training 39
4.5 Limitations of Study 39
4.6 Future Research Directions 40

5. CONCLUSION 40

6. ACKNOWLEDGEMENTS 41

7. REFERENCES 42

APPENDIX A: QUESTIONNAIRES 46
A.1. Background Questionnaire 46
1. A.1.1Summary of Background Questionnaire Responses 48
A.2. Exit Questionnaire 49

APPENDIX B: ADDITIONAL DETAILS RELATING TO GAME-BASED TRAINING 51
B.1. VBS2 Scenarios 52
2. B.1.1 ..... Practice Session 52
3. B.1.2 ..... Mission 1 53
4. B.1.3 ..... Mission 2 54
5. B.1.4 ..... Mission 3 54
6. B.1.5 ..... Mission 4 56

APPENDIX C: KNOWLEDGE-BASED ASSESSMENT 57

APPENDIX D: ANALYSIS OF TTCP GUIDEX THREATS TO A GOOD DEFENCE EXPERIMENT 59

List of Tables

Table 1: Summary of participants and assessors involved in study 10
Table 2: Section attack assessment criteria. The rating scale used was 0 = Not attempted/undertaken, 1 = Very Poor, 2 = Poor, 3 = Satisfactory, 4 = Good, 5 = Very Good12
Table 3: List and description of items removed from the game-based training data 13
Table 4: Schedule of activities for each day of the study 14
Table 5: Summary of Mann-Whitney testing of behaviour differences between sections ... 23
Table 6: Mean pre-training scores for each behaviour category using a competency-based rating scale. The numbers in parentheses represent the percentage of assessment ratings corresponding to a competent performance..... 23
Table 7: Mean scores for each behaviour across missions during game-based training..... 25
Table 8: Results of Wilcoxon Signed Ranks tests for pre- and post-training scores 28
Table 9: Cronbach's alpha values for each section's pre- and post-training scores 29
Table 10: Mean pre- and post-training scores for each behaviour category using a competency-based rating scale. The numbers in parentheses represent the percentage of assessment ratings corresponding to a competent performance..... 30
Table 11: Summary of exit questionnaire data shown as frequency of responses for each activity..... 33
Table 12: Summary of background questionnaire data (not included in body of report) ... 48
Table 13: Hardware and software components used during game-based training..... 51
Table 14: Running order of game-based training missions 52
Table 15: Frequency of ratings of 1 (as percentage of the total), χ^2 and p values (df =1) for pre- and post-training scores for both section. 57
Table 16: Frequency of ratings of 1 (as percentage of the total), χ^2 and p values (df =1) for pre- and post-training scores after removing 10 items from assessment criteria. 58
Table 17: List of GUIDEx threats to a good experiment and strategies used to mitigate them in the current study 59

List of Figures

Figure 1: Structure of an infantry section 7
Figure 2: Some of the participants with F88 Steyr (standing) and F89 Minimi (prone) 11
Figure 3: Some of the participants undertaking game-based training 16
Figure 4: Frequency of computer game-playing for each section 20
Figure 5: Self-reported confidence in conducting section attack procedures 21
Figure 6: Mean scores on pre-training assessment for each section 22
Figure 7: Mean scores for each behaviour on pre-training assessment for both sections 22
Figure 8: Mean scores of section performance for each mission during game-based training 24
Figure 9: Hit ratio data (as percentage) for BLUEFOR and OPFOR during game-based training. The assessor's ratings are plotted on the secondary axis. 26
Figure 10: Mean scores for pre- and post-training assessments for both sections 27
Figure 11: Frequency distributions of changes in score from pre- to post-training for each section 28
Figure 12: Difference between mean pre- and post-training scores for each behaviour category 29
Figure 13: Mean pre- and post-training scores for both sections, comparison of blind and non-blind assessors 30
Figure 14: Difference between pre- and post-training performance for each behaviour using a knowledge-based rating scale. 31
Figure 15. Difference between pre- and post-training performance for each behaviour using a knowledge-based rating scale after removing 10 items from assessment criteria. 32
Figure 16: BLUEFOR avatar (left) and OPFOR avatar (right) 51
Figure 17: Map of the island used in the practice session 52
Figure 18: Screenshot of the checkpoint in Mission 1 53
Figure 19: Map of Mission 1 53
Figure 20: Screenshot of the OPFOR location in Mission 2 54
Figure 21: Map of Mission 2 54
Figure 22: Screenshot of the terrain in Mission 3 55
Figure 23: Map of Mission 3 55
Figure 24: Screenshot of the terrain in Mission 4 56
Figure 25: Map of Mission 4 56
Figure 26: Comparison of average scores for pre- and post-training assessments using a knowledge-based rating scale. 57

This page is intentionally blank

Glossary

AAR	After Action Review
ADF	Australian Defence Force
AI	Artificial Intelligence
BLUEFOR	Blue Force
CBT	Computer-Based training
Comd	Commander
COTS	Commercial Off The Shelf
DSTO	Defence Science and Technology Organisation
EK	Enemy Killed
EW	Enemy Wounded
FPS	First Person Shooter
GUIDEx	Guide for Understanding and Implementing Defense Experimentation
HR	Hit Ratio
IET	Initial Employment Training
NHMRC	National Health and Medical Research Council
OPFOR	Opposing Force
p	Probability
RF	Rounds Fired
SD	Standard Deviation
SME	Subject Matter Expert
SOI	School of Infantry
TESS	Tactical Engagement Simulation System
TTCP	The Technical Cooperation Program
TTP	Tactics, techniques, and procedures
VBS2	Virtual Battlespace 2
WO2	Warrant Officer Class Two

This page is intentionally blank

1. Introduction

1.1 Background

Computer games are increasingly being used by military forces to supplement traditional methods of training instruction. The potential benefits of game-based training have been documented by several authors (e.g., Belanich, Mullin, & Dressel, 2004; Caspian Learning, 2008; Fournier, Lapointe, Munteanu, Edmond & Kondrovata, 2011), including the ability to conduct more training in less time, the potential to reduce the safety and logistical overheads associated with traditional training, the ability to provide personnel with a greater variety of training experiences, and the ability for personnel to train anywhere and anytime. In addition, game-based training is perceived to be a cost-effective solution to reducing military training budgets and maximising learning experiences and outcomes from live exercises (Prensky, 2001; Summers, 2012).

Historically, military forces have been using games for training and mission preparation for centuries (Smith, 2010). With the emergence of video games in the 1970s, these began to be adopted by military training institutions (Temby, 2005). As computing technology has increased over time, so too has the level of game sophistication and interest from educators in using them as training devices. Currently, computer game technologies are used for training by a variety of military forces, including Australia, the USA, Canada, and the UK (Bell, Jarmasz, & Nelson, 2011; Padilla, 2012; Ratwani, Orvis, & Knerr, 2010; Whitney, Fidock, & Ferguson, 2012).

The investment in game-based technology has been expensive. For instance, when the US Army's game-based training program was initially announced, the budget was $50 million (Atherton & Baxter, 2009; Robson, 2008). Given the substantial investment in these technologies, and the need to optimise their use, including preventing any undesirable outcomes (e.g., negative learning), researchers have emphasised the need to evaluate their effectiveness (e.g., Belanich et al., 2004; Knerr, 2006; Topolski et al, 2010; Whitney, Temby, and Stephens, in press).

This report documents the outcomes of a DSTO study examining the effectiveness of a desktop computer game to train teams of dismounted soldiers in infantry tactics, techniques and procedures (TTPs). The study was commissioned by the Australian Army as part of a request to evaluate the efficacy of simulation technologies, including desktop computer games, for Army training purposes. This study was sponsored by Training Command – Army[1] and conducted by scientists from the DSTO's Land Operations Division working on the Training and Preparedness (ARM 07/163) task.

[1] Subsequent to the conduct of this study, Training Command – Army became part of Headquarters Forces Command Army.

1.2 Previous Research

In the following paragraphs, we provide a brief review of studies that have examined the effectiveness of game-based training with military populations. The review is restricted to studies involving dismounted soldiers because of the focus of the current study on this population. As noted by Whitney, Temby, and Stephens (2008), there are only a few published studies that have evaluated the effectiveness of desktop computer games for training dismounted soldiers. Consequently, what follows is a reasonably comprehensive summary of published work in this area.

In this review, studies examining the effectiveness of computer-based instruction methods (e.g., 'e-learning') with military personnel (e.g., Ricci, Salas, & Cannon-Bowers, 1996; Parchman, Ellis, Christinaz, & Vogel, 2000) have been excluded. This is because these studies do not involve the use of first-person perspective computer games that typically characterise the field of studies on game-based training, nor do they examine the training of field-based, dismounted soldier tasks. Rather, they involve the use of self-paced learning modules, often in the form of interactive PowerPoint slides, with little to no involvement from instructors. Consequently, while studies have found computer-based instruction to be effective (see De Freitas, 2006; Hays, 2005; and Kulik, 1994, for reviews of studies in this area), they are not comparable for the above reasons. Finally, we do not review game-based studies that only examine soldier performance in the virtual environment (e.g., Beal & Christ, 2004), or compare soldiers' perceptions of their performance in the live environment (the "real world") and virtual environment (e.g., Beal, 2005; 2007; 2009; Lewis, 2005; Morrison, Barlow, Bethel, & Clothier, 2005; Rencrantz, 2003). This is because these studies do not examine whether these skills transfer to real world tasks or do not provide corroborating evidence, and hence have not validated their findings.

In a study with British Army personnel, Pennell (2003) investigated the effectiveness of the desktop computer game Half-Life to train building clearance procedures. The study involved 16 military personnel undertaking either traditional training or a blend of game-based training and traditional training. The performance of both groups was assessed by subject matter experts (SMEs) after training during a subsequent field activity. The findings suggested that the group that undertook game-based training performed better on the post-training activity; specifically they were deemed to be better at decision-making, time to clear rooms, and use of grenades. While the findings were encouraging, there were several shortcomings in the study that limited the strength of the conclusions that were drawn, such as the use of subjective rather than objective data, and the small sample size; performance was assessed at the team rather than individual level, reducing the sample size to four 4-man teams[2].

In a study with US Army officers, Nolan and Jones (2005) examined the effectiveness of the game Delta Force: Black Hawk Down for training tactical skills. The study involved 41 officers who were allocated to either an experimental group, which received sixteen hours of game-

[2] While assessing performance at team level rather than individual level is appropriate for studies examining team-based training, in this study it is considered a limitation as it reduced the sample size to a number which Pennell (2003) considered too small for statistical analysis.

based training, or a control group. Both groups received coursework instruction as part of their normal training prior to the experiment[3]. One methodological strength of the study is that it looked for potentially confounding variables between the groups, by measuring factors such as 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, they concluded that any performance differences between the groups were more likely to have resulted from training differences rather than any differences in skill and experience.

Following training, the participants took part in a field exercise in which their actions on contact were assessed by SMEs. Although the assessment is not described in detail, the SME suggested that the performance of the game-based training group was better than the group that completed normal coursework, in terms of their movement, awareness and planning. While Nolan and Jones (2005) do not acknowledge this issue, this outcome may simply indicate that game-based training is better than no training at all. It does not provide any evidence of the effectiveness of game-based training relative to traditional training methods (which would have required the control group to receive 16 additional hours of coursework). This highlights the need for appropriate comparison groups when evaluating the effectiveness of game-based training. Another limitation is that the assessors were aware of which participants took part in game-based training. Finally, over 60% of the study participants had prior military experience, and 5% had operational experience. This prior experience may have reduced the likelihood of detecting measurable performance improvements in the study. Consequently, while the study findings provide some support for the effectiveness of game-based training, there are a number of potential confounds which limit the strength of the findings.

In a series of studies with US military personnel, Wiederhold (2005) examined the effectiveness of an unspecified computer game to train urban TTPs, including building clearance procedures. The procedure, findings, and conclusions were similar across studies; for brevity, only one study is described here. The participants (n=210) were allocated to an experimental group (n=90, that received approximately 15 minutes of game-based training) or a control group (n=120, that received no specific training). Participants in the experimental group practiced navigating and conducting building clearance drills in a virtual environment, or "shoot house". This consisted of a virtual building or village, containing friendly, hostile and neutral personnel, and items of interest. Both groups then completed an exercise involving building clearance drills inside a real village and shoot house. The performance of the groups was assessed by SMEs on a number of performance measures, including time taken to clear rooms. The study reported that the game-based group outperformed the control group in the field exercise, including quicker decision-making, better awareness of surroundings, and better organised movement. Based on this data, the researcher concluded that the findings provided evidence for the effectiveness of game-based training. However, as with the previous studies, there were aspects of the methodology that limit the conclusions that can be drawn. Firstly, no baseline performance data was collected, which precludes any assessment of changes in pre-and post-training performance. Secondly, the limited amount of training time on the game (approximately 15 minutes), together with the significant previous

[3] The details of the coursework are not provided in Nolan and Jones' (2005) report.

military experience of the participants raises doubts about the level of training benefit conferred by the game. Thirdly, the virtual shoot house in the game was a direct replica of the real shoot house, so it is likely that the game-based training provided familiarisation with the live environment, which was not available to the participants in the control group. Therefore, the outcomes simply show that the game-based training was better than no training, especially when this training involved conducting identical tasks to that subsequently assessed in the live environment. Finally, the participants were assessed on skills in the field exercise that were not capable of being conducted in the game's virtual environment, such as the ability to search suspects. As a consequence, the evidence for the game's effectiveness in this study is highly confounded by these other factors.

In a study with US Army personnel, Kneuper (2006) examined the effectiveness of the game Delta Force: Black Hawk Down to train tactical skills to officer cadets. In a comprehensive study conducted over five months, the participants were allocated using stratified random sampling on the basis of gender, ethnicity, prior computer game experience, and academic ability into six different groups. The use of stratified random sampling to avoid potential confounds is a strength of the study. Following allocation to groups, each group received a mix of different amounts of game-based training (0%, 15%, 30%, 45%, 60%, and 75%) relative to the amount of live training received. The total amount of tactical training received during the course was 20 hours. At the completion of training, each participant was rated by SMEs on 16 behavioural dimensions during a field exercise. The use of SMEs to assess the participants is a strength of this study, however, some of the behavioural dimensions, such as physical fitness and military bearing (represented by demeanour and tone), appear incapable of being trained with games.

Based on mean scores in this study (Kneuper, 2006), there were no statistically significant differences between the performances of any of the groups. However, because the 45% group had the highest mean scores across the 16 behavioural dimensions, Kneuper (2006) concluded the optimum mix of training was 45% simulation and 55% live. This conclusion is ambitious, and cannot be sustained on the available evidence. In addition, there are several methodological and conceptual weaknesses with this study. These weaknesses include: (1) the use of a 3-point rating scale (unsatisfactory, satisfactory, excellent) to assess performance, which was not sensitive to detecting small differences in performance, (2) the assessment of behaviours that cannot be trained in a virtual environment, such as physical fitness and military bearing, and (3) the tendency for the assessors to give 'satisfactory' ratings to participants, all of which could plausibly account for the lack of statistically significant differences between the groups and prevented a rigorous evaluation of the game's effectiveness. Overall, while some of the methodology was very good, the study was let down by these limitations.

Proctor and Woodman (2007, see also Woodman, 2006) compared the performance of two groups of US Marine Corp trainees undertaking close combat training: one group (n=16) completed traditional training (involving 3 hours of lectures and 3 hours of shoot house walkthroughs) and the other group (n=16) completed a modified training method using the game Close Combat: First to Fight (consisting of 2 hours of classroom lectures, 2 hours of game-based missions, and 2 hours of shoot house walkthroughs). Both groups then completed a post-training test on which their performance in close combat procedures was assessed.

There was no significant difference in the performance of the two groups at the end of training; however the researchers concluded that the results supported the efficacy of game-based training. While the study employed some useful control measures to increase the reliability of the results (e.g., blind assessment methods), a number of methodological issues were evident in the study. For example, 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. In addition, there was no pre-test prior to training, so the researchers were unable to measures changes in performance following training. Finally, while levels of computer game use were assessed in the experimental group, they were not assessed in the control group, raising the possibility that the two groups had differing levels of computer game experience; a potential confound. Overall, the findings indicate that there was no significant difference between the training methods; due to the above methodological factors, little can be deduced about the effectiveness of either training method.

In a study with infantry soldiers, van der Hulst, Muller, Besselink, Coetsier, and Roos (2008) examined the use of Virtual Battlespace 2 (VBS2) for training infantry section attack. In their study, 9 Dutch Army personnel taking part in a training course participated in an infantry section attack activity executed in VBS2 under the guidance of an instructor. The authors reported that the students and instructor believed that performance had improved following training, although no other empirical evidence is included to support this. Following game-based training, the students completed a series of field exercises. Van der Hulst et al. state that in the live environment, there was an initial decrease in performance, but after one or two repetitions of an exercise, performance levels returned to those seen during game-based training. While this study demonstrates that VBS2 has been used for training infantry section attack, it is difficult to draw any conclusions about its effectiveness, given (1) there was no control group or pre-training assessment, and (2) the evidence provided in support of transfer of training and performance improvement during training is anecdotal, and the eventual performance improvement in the field could simply have arisen due to learning effects from repeating the activity.

In summary, it appears that while previous studies examining game-based training with small teams of dismounted soldiers have demonstrated some methodological rigour, they have also been hampered by methodological weaknesses including: lack of control (or comparison) groups, lack of baseline assessments, small sample sizes, not using blind assessment protocols, inability to isolate the effect of game-based training and use of unreliable performance measures. As a result of these shortcomings, there are few conclusions that can reliably be drawn from these studies; at best it can be concluded that game-based training does not appear to cause any negative learning effects on dismounted soldiers. This conclusion is consistent with the findings from more general reviews of game-based studies with other populations (see De Freitas, 2006; Hays, 2005). Consequently, we believe that more rigorous studies are needed before definitive conclusions can be drawn. While the current study was driven mainly by a request from our military customer to investigate the effectiveness of game-based technologies for military training, through addressing this request we were also able to address methodological limitations in (and extend) previous research in this area.

1.3 The Current Study

The current study was conducted as part of a larger research program in DSTO's Land Operations Division, which is investigating the potential benefits of new technologies for individual and collective training in the Australian Army. As previously stated, the study addressed a request from the sponsor to evaluate the effectiveness of game-based training. The current study addressed this requirement by examining the effectiveness of the desktop computer game VBS2 to train teams in dismounted infantry procedures. VBS2 was selected as it is currently in use within the Australian Army, is promoted by the manufacturer as effective for training dismounted combatants, and has previously been used for infantry training by other military forces (van der Hulst et al., 2008). The study was conducted over five days from 1-5 September, 2008 at the School of Infantry (SOI) in Singleton, New South Wales, Australia.

1.3.1 Research Questions and Objectives

The two main research questions addressed by the study were:

1. Can game-based training be effective for teaching combat tactics, techniques, and procedures (TTPs) to novice dismounted infantry soldiers?
2. Do the skills and procedures trained in virtual environments using desktop computer games transfer to the live environment (i.e., real world)?

The specific research objectives were to:

1. Compare training outcomes from game-based and field-based training.
2. Evaluate the effectiveness of VBS2 for training novice infantry soldiers in section attack procedures.
3. Evaluate a methodology for evaluating game-based training.
4. Identify the implications of the findings and provide recommendations for future research into game-based training.

The following sections include a brief description of an infantry section, section attack procedures, and VBS2 to provide additional background information for the reader before outlining the methodology employed.

1.3.2 The Infantry Section

An infantry section in the Australian Army comprises nine men[4], divided into three groups: the command group, the assault group, and the gun group (Barlow, Morrison, Luck, & Dickie, 2004). Each group has a specific function and the members of each group have specific roles and responsibilities (Guille & French, 2004). The structure of an infantry section is shown in Figure 1. The command group comprises the section commander, who is responsible for

[4] At the time the study was conducted, entry to the Royal Australian Infantry Corps was restricted to males only; hence gender-specific language is used in descriptions of an infantry section and participants in this study.

commanding the entire section, and two scouts, whose role is to gather and report information to the commander. The assault group comprises a gunner and two riflemen, and their function is to assault enemy positions. The gun group is comprised of the second-in-command (2IC), a gunner, and a rifleman (known as #2 on the gun, although he does not carry a machine gun); the function of the gun group is to provide fire support during a section attack (Colton, 2008; Hobbs & Mouzakis, 2001; personal communication with Army Warrant Officer, 2008).

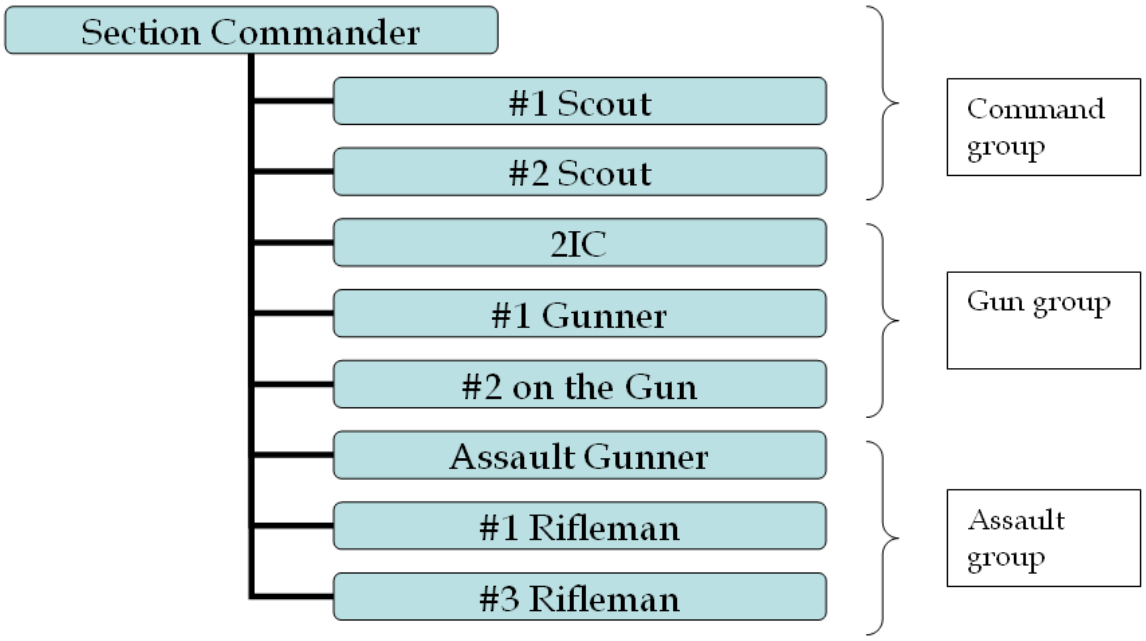

Figure 1: Structure of an infantry section

1.3.3 Section Attack

The section attack is a military team activity where the aim is to destroy the enemy or capture terrain using a combination of coordinated fire and movement (Curtis & Hobbs, 1997). A section attack has three main phases: preparation, contact with the enemy, and reorganisation after contact. In the preparation phase, the section commander delivers orders to the rest of the section, and the section moves in formation to the enemy location. When contact with the enemy occurs, the section will move and fire as directed by the section commander, in order to achieve a particular objective, typically to eliminate the enemy or gain control of the location. Finally, in the reorganisation phase, the section regroups, and checks things such as the amount of ammunition remaining and the status of any casualties. In the Australian Army, a section attack is typically trained in a 'crawl, walk, run' manner, where individual phases of the attack are rehearsed or walked through at a slower pace than normal, with the intensity gradually increasing until it approaches realistic levels.

The section attack involves both physical (e.g., fire and movement) and cognitive skills (e.g., knowledge of procedures, choosing fire support positions, maintaining awareness of other

team members' positions). While the physical aspects can be represented in the virtual environment, their execution uses very different skills than those used in the live environment (e.g., using a keystroke to reload a weapon as opposed to changing the magazine). However, in general the section attack is largely assessed on the performance of the section with a focus on cognitive skills rather than individual physical skills. The Australian Army currently trains section attack using a combination of classroom lessons (i.e., to cover the theory of a section attack) and field-based activities (i.e., to practice the fire and movement procedures as a section). While the Australian Army does not currently use computer-based training methods to teach section attack, the study focused on infantry section attack for the following reasons: (1) the section attack is a fundamental infantry activity and (2) the section attack involves skills that can be executed and assessed in live and virtual environments[5].

1.3.4 Virtual Battlespace 2

VBS2 is a first-person perspective simulation developed by Bohemia Interactive Australia (BIA). VBS2 provides a dynamic virtual environment that allows users to participate in training scenarios (or 'missions') from a first-person 3D perspective and a 2D tactical commander view.

VBS2 is currently used by the ADF to support mission rehearsal activities, such as terrain familiarisation, during pre-deployment training, and to support driver training for convoy vehicles and armoured vehicles (Law, 2008). VBS2 includes an After Action Review (AAR) capability that provides a range of summary performance statistics for each player and mission, such as numbers of rounds fired, numbers of enemy killed or wounded, numbers of friendlies killed or wounded, and the hit ratio (which is the ratio of the number of the opposing side killed or wounded to the number of rounds fired).

In their studies, Barlow and colleagues (Barlow, et al. 2004; Morrison & Barlow, 2004) have used hit ratio data as a measure of infantry section effectiveness. However, Barlow et al. (2004) did not comment on whether this data is a useful performance measure in the context of their study. More recently, the utility of automated performance measures was examined by Hussain and Feurzeig (2008), who reviewed the type of information provided by the AAR capabilities of various computer games, and concluded that these tools are currently insufficient for accurate assessments of individual and team performance. However, they did not provide any experimental evidence to support this. Although it was not a formal research objective, the presence of the AAR capability provided the opportunity to collect automated performance data, and compare it with SME ratings to examine if Hussain and Feurzeig's conclusions are supported.

[5] Earlier in this report we criticised Kneuper (2006) for assessment of skills that could not be trained through a computer game. We believe that our study differs in that the skills our participants were assessed on are capable of being represented at a rudimentary fashion (or better) in VBS2, whereas in his study several skills could not be represented at all (such as military bearing), even at low levels of fidelity, in the virtual environment.

2. Method

2.1 Study Design and Ethical Approval

A 2 x 2 factorial (groups x training method), pre- and post-test experimental design was employed. Each section[6] completed a pre-training test to assess their baseline performance in conducting a section attack. One section then undertook eight hours of section attack training using VBS2 (i.e., the game-based section); the other section received eight hours of traditional instruction in section attack procedures (i.e., the field-based section). At the end of training, both sections completed three section attacks in the field to allow the amount of training transfer to be measured reliably.

The study received ethics approval from DSTO's internal human research ethics panel prior to data collection and was conducted in accordance with the ethical principles outlined in the National Statement on Ethical Conduct in Human Research (NHMRC, 2007).

2.2 Participants and Assessors

Twenty two soldiers from the Australian Army took part in the study (Mean age = 22 years, SD = 5 years). Eighteen of these participants (16 Privates, and 2 Corporals (section commanders), comprised the game-based and field-based training sections. The remaining four participants (2 Lance Corporals, and 2 Privates) played the role of Opposing Force (OPFOR) during the game-based and field-based activities. The two Section Commanders were training staff from the School of Infantry (SOI) and both were qualified to train section attack. The other sixteen members had recently completed Army Recruit Training and therefore had a basic knowledge of infantry TTPs. These trainees were an ideal sample for the study because (1) they would be more sensitive to learning effects compared with more experienced soldiers and (2) they had a basic level of military experience sufficient to enable them to undertake section attack training (compared to participants with no military experience). While the OPFOR participants were not specifically trained for their role, they were experienced in infantry section attack procedures, and were given guidance by the researchers and military SME in how they should carry out their role to support the study objectives. In addition, while some data were collected from the OPFOR participants (see Sections 2.5.8 and 3.4.4), the results were not included in analyses unless otherwise stated.

Five Australian Army personnel (3 Warrant Officers and 2 Sergeants) were employed as assessors in the study; their role was to assess each section's performance at section attack before, during and after training. One assessor evaluated the section's performance during game-based training and another assessor evaluated the other section's performance during field-based training. Four assessors (including the one who assessed the field-based training) were involved in the assessments conducted before and after training; two of these assessors

[6] For the sake of clarity, the two groups are referred to as the 'sections' throughout the report. This is to avoid potential confusion with the three groups (command group, gun group, and assault group) that make up an infantry section.

were blind to the training method undertaken by each section[7]. A summary of the participants and assessors is outlined in Table 1.

Table 1: Summary of participants and assessors involved in study

Participants	Assessors
1 x 9-man section (game-based section) 1 x 9-man section (field-based section) 4 x OPFOR (game-based training, field training, and pre- and post- assessments) Total = 22	1 x assessor (game-based training) 1 x assessor (field-based training and pre- and post- assessments) 3 x assessors (pre- and post- assessments) Total = 5

2.3 Equipment

2.3.1 Computer Hardware and Software

Thirteen networked computers (9 Blue Force (BLUEFOR), 3 OPFOR and 1 observer station) were used for the game-based training. The computer specifications all exceeded the minimum system requirements listed on the VBS2 website at the time the study was conducted (BIA, 2008). The hardware and software components used for game-based training are shown in Table 13 in Appendix A. Headphones were worn by participants to monitor in-game audio cues[8]. Because participants were located close together, voice communication was adequate for the exchange of information between section members.

2.3.2 Weapon Systems

The following weapon systems and equipment were used in the field activities:

- Tactical Engagement Simulation System (TESS): the in-service ADF version of TESS equipment was worn by participants during the section attack missions in the field to provide simulated weapons effects realism associated with force-on-force engagements.
- F88 Steyr and F89 Minimi: each section member was allocated either an F88 or F89 weapon for the duration of the study. The F88 Steyr is the assault rifle used by infantry soldiers in the Australian Army. The F89 Minimi is a machine gun which is used to provide suppressive and covering fire. There are two F89 Minimis and seven F88 Steyrs in an infantry section. A picture of section members with these weapons is shown in Figure 2.
- Blank ammunition for the F88 and F89 weapons was used to provide additional levels of realism to the section attacks conducted in the field.

[7] Every effort was made by the study team and SOI staff in the planning stages to have a totally 'blind' assessment of each section's performance during the post-training assessments. However, staff shortages during the study meant that only two of the four assessors were fully blind. The impact of this is discussed in Section 3.4.1.

[8] The in-game sound effects were produced in stereo (two-channels) but were not 360°.

Figure 2: Some of the participants with F88 Steyr (standing) and F89 Minimi (prone)

2.4 Measures

2.4.1 Background Questionnaire

A background questionnaire was used to collect demographic data on each participant. The questionnaire contained 18 items relating to participants' age, length of military service, computer game experience, and self-ratings of confidence in conducting section attacks. This data was collected and used to allocate participants into the two sections and to control for any potential confound effects in subsequent data analysis; such methods have been used in previous studies. A copy of the background questionnaire is contained in Appendix A.1.

2.4.2 Section Attack Assessment Criteria

The criteria and rating scale used for assessing section attack are listed in Table 2. The criteria are the same used by the SOI for assessing section attack. The criteria were compiled into a 35-item checklist which was used by the assessors for all missions. The criteria are grouped under the categories of communication, leadership, coordination, supporting behaviour and general points, which are consistent with critical team dimensions in the team training literature (e.g., Smith-Jentsch, Zeisig, Acton, & McPherson, 1998). A six-point rating scale (ranging from 0 to 5) was used to provide more granularity and sensitivity to small performance changes without being too difficult for the assessors to use. The rating scale and interpretation of each rating score was discussed between the study team and the assessors prior to the study, in order to ensure a common understanding of the assessment requirements.

Much of the Australian Army's assessment of training utilises a competency-based assessment, where performance is assessed as either 'competent' or 'not yet competent'. Although not formally stated as a research objective, the checklist was also used for competency-based and knowledge-based assessments, as discussed further in Sections 3.2.1, 3.4.2, 3.4.3 and Appendix C, to see if the outcomes were sensitive to different assessment methods.

Table 2: Section attack assessment criteria. The rating scale used was 0 = Not attempted/undertaken, 1 = Very Poor, 2 = Poor, 3 = Satisfactory, 4 = Good, 5 = Very Good

Communication
Warning Order Issued
Section Attack Orders
Fire Control Orders During Assault
Passage of Info - During Assault
Passage of Info - During Reorganisation
Information - Timely, Relevant, Clear
General Points
Movement Aggressive
Weapon Stoppages Cleared
Muzzle Awareness
Fire Positions Suitable
Enemy Position Covered
Close Assault Drills Conducted
All Enemy Identified
Position Neutralised
Limit of Exploitation Achieved
Supporting Behaviour
Reaction of Troops to Directions
Covering Fire Effectiveness
Fire Support Effectiveness
Effective Fire on Enemy
Pits Cleared Effectively
Casualties Treated During Reorganisation
Leadership
Section Comd Aware of Position of Troops
Section Comd Control
Group Comd Aware of Position of Troops
Group Comd Control
Assault Formations Suitable
Control During Reorganisation
Coordination
Movement into Assault Formations
Position of Fire Support Group
Control of Fire Support Group
Spacings Suitable
Bounds Appropriate
Groups Work in Teams/Pairs
Fire and Movement Effective
Casualties & Ammo Checked During Reorganisation

It was noted that current limitations with VBS2 restricted the participants' ability to execute certain skills required when conducting a section attack. In addition, there were items that could be omitted on the basis that they could be equally well executed regardless of the means of training (for example, issuing a warning order is simply a verbal or written command). Consequently, the assessment criteria in Table 2 were reviewed by the assessor involved in game-based training and ten items were removed from the assessment criteria (see Table 3 for a list of the items and rationale for their removal) use for game-based training.

Table 3: List and description of items removed from the game-based training data

Item	Rationale for removing items
Warning Order Issued	Issuing a warning order is a simple verbal command carried out by the section commander. It was not deemed necessary to assess this skill because of the ease of conducting it.
Passage of Info- During Reorganisation	The reorganisation procedure is difficult to conduct in the virtual environment due to the precise movements required.
Control During Reorganisation	The reorganisation procedure is difficult to conduct in the virtual environment due to the precise movements required.
Casualties and Ammo Checked During Reorganisation	Checking for casualties and ammunition is difficult to conduct in the virtual environment due to the precise movements required.
Pits Cleared Effectively	Clearing pits is a task that is difficult to conduct in the virtual environment because of the precise movements required.
Casualties Treated During Reorganisation	In the virtual environment, treatment of virtual casualties can be achieved with a single keystroke, therefore it is not comparable with the live environment.
Movement Aggressive	The aggression of movement is difficult to assess in the virtual environment; it reflects a combination of speed, intensity, and determination which are hard to judge from observing the avatar in the virtual environment.
Weapon Stoppages Cleared	Weapon stoppages do not occur in the virtual environment.
Muzzle Awareness	It is difficult to assess whether trainees are maintaining awareness of their weapon's muzzle position in the virtual environment.
Limit of Exploitation Achieved	There was no opportunity to assess this item in the virtual environment. When the mission objective was achieved (i.e., all enemy destroyed), the mission was terminated. At this point, the section commander briefed the section members on what would normally be done at this point, including confirming the limit of exploitation had been achieved.

2.4.3 AAR Data

The number of rounds fired, friendlies killed or wounded, enemies killed or wounded, and hit ratio data for each mission were automatically recorded by VBS2's AAR capability. This data was subsequently used to assess the game-based training sections performance during the game-based training sessions.

2.4.4 Exit Questionnaire

The exit questionnaire was used to collect participant ratings and comments regarding their perceptions of game-based training, including the ease with which they could execute certain skills in the virtual environment. The questionnaire was developed and modified from questionnaires used in previous studies of dismounted soldiers (e.g., Nolan & Jones, 2005). The questionnaire contained 11 items and asked participants to respond on the following scale: *1 = very easy, 2 = somewhat easy, 3 = somewhat difficult, 4 = very difficult.* With regard to an item relating to training effectiveness, participants were asked to respond on the following scale: *1 = not at all effective, 2 = somewhat effective, 3 = moderately effective, 4 = very effective*. The questionnaire was administered to the game-based section and OPFOR participants on the final day of the study. A copy of the questionnaire is included in Appendix A.2.

2.5 Procedure

The study was conducted in accordance with the procedures described in the following sections and the schedule in Table 4. The total training time of 8 hours was based on the approximate time allocated to section attack training during Initial Employment Training (IET) in the Australian Army, and was considered adequate by the training staff at the SOI for conferring a measurable training benefit on the participants.

Table 4: Schedule of activities for each day of the study

Day	Activity
1	Introductory Brief by study team
	Completion of Background Questionnaire
	Participant Allocation to Sections
	Theory Lessons on Section Attack (2 x 45 minutes)
	Pre-Training Assessment: one section attack in field per section
2	Section Attack Training (4 hours)
	- Game-based training for one section
	- Field-based training for other section
3	Section Attack Training (4 hours): continuation of training
	- Game-based training for one section
	- Field-based training for other section
	Post-Training Assessment: 1 mission per section
4	Post-Training Assessment: 2 missions per section
5	Completion of Exit Questionnaire
	Debrief Session

2.5.1 Introductory Brief and Background Questionnaire

On the first day of the study, the participants and assessors were given an overview of the study background and objectives by the study team. Participants were given study information sheets, and were invited to ask any questions they had about the study. Following this, the participants signed consent forms and completed the background questionnaire.

2.5.2 Allocation to Sections

After completing the background questionnaire, the participants were divided into two 9-man sections: a game-based training section and a field-based training section. To obtain compositions that were comparable, the participants were allocated into sections using stratified random sampling on the basis of their self-reported computer game experience. The composition of each section was then checked to ensure the two sections contained participants with equivalent amounts of military experience and previous training in section attack.

2.5.3 Theory Lessons

After the participants were allocated into sections, they received two theory lessons on section attack, each lasting approximately 45 minutes. The lessons covered the phases of the section attack and the different types of section formations. The lessons were delivered by the two section commanders using a combination of PowerPoint slides and whiteboard demonstrations. At the end of each lesson, participants were questioned by the section commanders on their knowledge of the material covered to confirm learning objectives had been met. The purpose of the lessons was to ensure that all participants had been exposed to the same information and to reinforce their basic knowledge of section attack procedures. The lessons also helped to prepare the participants for the training they would be undertaking for the remainder of the study.

2.5.4 Pre-Training Assessment

Following the theory lessons, both sections undertook a pre-training assessment. This assessment was undertaken to obtain a baseline measure of each section's ability to conduct a section attack. The assessment was conducted in the Close Training Area in the Singleton Military Area, which provided suitable terrain for conducting section attack activities. A single assessment was conducted because additional assessments could have provided learning opportunities prior to the training sessions being undertaken (which could have potentially contaminated the rating data due to practice effects). The assessment was conducted by four assessors using the checklist in Table 2. One assessor was positioned immediately behind each of the command, gun and assault groups. Consequently, the ratings by each of these three assessors largely reflect the performance of that group, although it also takes into account the performance of the section commander and the rest of the section. The fourth assessor was positioned with the OPFOR and provided an overall assessment of the section. The assessment was conducted in real-time for all phases of the attack. At the completion of the assessment, each section commander provided feedback to his section to reinforce learning points and areas for improvement. The role of the assessor was to assess the performance of each section; no feedback was provided by assessors to either section (in the pre- and post-training assessments) to prevent the possibility that one section received more assistance than the other. It is also worth noting that each section was prevented from observing the performance of the other section while conducting the section attack; this was achieved by having the other section in a different location at the time. This was done in order to prevent observational learning.

2.5.5 Section Attack Training

Section attack training for each section was conducted immediately after the pre-training assessment. The training sessions were conducted in parallel to ensure equivalent amounts of training time for both sections. While the field-based section did not conduct the familiarisation training that was necessary for the game-based section, when taking into account the time required for the field-based section to move to and from field locations, the amount of time actually spent training on section attack for both sections was not significantly different (i.e., in the order of a few minutes). A detailed description of each training method is provided below.

2.5.5.1 Game-Based Training

Game-based training was conducted in the Computer Learning Facility at the SOI. The training was conducted with participants arranged in a linear formation as shown in Figure 3. Each participant was assigned to a computer station (with keyboard, mouse and monitor) corresponding to their allocated position in the section. Each computer was labelled with the section member's position to facilitate role identification for the assessor throughout training.

Due to the layout of the facility, the game-based section and OPFOR were co-located in the same room. The three OPFOR participants were physically separated from the game-based section using office partitions to maintain some tactical realism and prevent the OPFOR from being able to observe the actions of the game-based section. As the game-based section communicated verbally rather than through hand signals, the OPFOR could hear what they were saying. However, the OPFOR were given instructions on how to engage the BLUEFOR, and were supervised by the assessor to ensure that their actions supported the training objectives. While the game-based training section could overhear the OPFORs' communication, the OPFOR participants did not communicate much during training, and did not discuss in detail their position or intended approach. Moreover, the OPFOR participants would not have become aware of the BLUEFOR avatars' locations until they approached their position and came into view. Therefore, having OPFOR located in the same room as the section, while not ideal, is not likely to have affected the section's training outcome.

Figure 3: Some of the participants undertaking game-based training

2.5.5.1.1 Game Familiarisation

Game-based training commenced with an introductory session to allow participants to become familiar with the game controls and virtual environment. During the session, the participants individually completed two tutorials in VBS2; one on basic movement and another on how to use weapons in the game. Following this, the section practiced basic fire and movement skills as a team within the virtual environment. The familiarisation session lasted for three hours after which time the assessor deemed that all members of the game-

based section were sufficiently competent in the use of the game controls to move onto specific training[9].

2.5.5.1.2 Training Missions

Once the familiarisation session was completed, the game-based training proper began. The training was provided by the section commander under the supervision of the game-based training assessor. The role of the assessor was to assess the section's performance, ensure the OPFOR supported the training objectives, and provide additional feedback to the section. The section commander was instructed to conduct the training as he normally would within the time allowed and the constraint of having to use the computer game.

The training included the following activities:

- Theoretical instruction in the stages of the section attack
- Conduct of section attack missions with OPFOR including walkthrough methods using the real-time 2D view projected onto a screen visible to all section members
- Debrief on section attack performance by the assessor using the 2D and 3D AAR tools
- Debrief on team performance by the section commander including theoretical instruction in the stages of section attack and revision of basic infantry TTPs.

In each mission, the OPFOR were played by military personnel from the SOI, rather than using Artificial Intelligence (AI) controlled avatars. The decision to use real OPFOR was made for two reasons. Firstly, having the OPFOR played by humans added realism to the training and a sense of competition for the game-based section. Secondly, AI-controlled avatars in VBS2 have limited and sometimes unrealistic patterns of behaviour. For instance, as noted by BIA staff (Jarvis, 2008), programming limitations in the version of VBS2 used at the time of the study allowed AIs to walk through walls and other solid objects. In addition, the Army SME involved in the scenario development felt that the response of the AI entities to contact was unrealistic; for instance, when shots were fired, they remained in the open and did not take cover.

During training, the OPFOR players were instructed to engage the section when they came within range. However, given that the OPFOR players were more experienced soldiers than the section (excluding the section commander), with greater knowledge of tactics and in one case greater familiarity with VBS2[10], they were given guidance and constraints on their behaviour by the assessor to ensure they supported the training objectives and were not simply trying to outperform the section.

The missions were created using the VBS2 Mission Editor. In developing the missions, the study team and an Army SME (who served as the assessor in the game-based training) considered the following factors: type of terrain, number of OPFOR, insertion point, and point

[9] While no members of the game-based section had played VBS2 previously, 8 of the 9 members had played similar first-person shooter games previously and were familiar with the game controls, compared to 7 out of 9 in the field-based section.

[10] On the background questionnaire, one of the OPFOR participants reported having played VBS2 1-3 times. He also reported playing computer games daily.

of contact. All missions took place in the VBS2 fictional location of Sahraini. This area was representative of the terrain in the Singleton Training Area; it included moderately wooded sections with rolling hills and foliaged areas which provided adequate cover and concealment from enemy fire.

Once an appropriate location for each mission had been identified and selected, OPFOR and BLUEFOR avatars were added to the scenario. The insertion point varied across missions, but was typically several hundred metres from the OPFOR starting location. Each of the missions was slightly different in terms of the terrain, direction of approach, and location of OPFOR. Despite these changes, the assessor's opinion was that the overall level of mission difficulty did not vary during the training; this was done to allow an assessment of learning to be made over the missions. The missions were intended to be achievable but challenging. Each mission was recorded using VBS2's AAR function. Full descriptions of each mission including images and maps are provided in Appendix B.

2.5.5.2 Field-Based Training

Field-based training was conducted in and around the Singleton Training Area by the commander of the field-based section. The section commander was instructed to conduct the training as he would normally but within the time allowed. The field-based training method employed in the study was the same method currently used by the Australian Army. The training included the following activities:

- Theoretical instruction in the stages of the section attack
- Practice in the stages of the section attack using a walkthrough method
- Conduct of complete section attack with OPFOR
- Debrief on team performance by the section commander including theoretical instruction in the stages of section attack and revision of basic infantry TTPs.

In all, three complete section attack missions were conducted using this approach. This is less than the number of missions completed by the game-based training section. However, the amount of time each section spent on training was the same (8 hours)[11].

2.5.6 Within-Training Assessment

During training, the performance of the game-based section was assessed for each of the six missions by the assessor using the criteria in Table 2. Similarly, the performance of the field-based section was assessed during training by one of the assessors for three missions. These three missions were conducted towards the end of the training session. Earlier missions were conducted as walkthroughs rather than complete section attacks, therefore the assessor did not rate the section's performance for the earlier missions. Both sets of data were used to quantify the performance of each section during the training sessions and therefore assess the

[11] Giving each section equal training time assumes that a period of simulation-based training and a period of field-based training provide the same training benefit. This assumption may not be correct. For instance, it may require greater (or lesser) amounts of simulation-based training in order to achieve the same outcome as field-based training. It is beyond the scope of this report to resolve this issue.

extent to which section attack skills were being learned by the participants in each training condition.

2.5.7 Post-Training Assessment

Following training, both sections moved to the field for the post-training assessment. Each section completed three section attacks in order to obtain a representative measure of their performance after training. Each of the scenarios was slightly different in terms of the terrain, direction of approach, and location of OPFOR. These modifications were included to reduce the likelihood of participants becoming too familiar with the terrain, which could have occurred if all three scenarios were identical. The post-training assessment was assessed in the same way as the pre-training assessment, using the criteria in Table 2 and with the assessors positioned in the same locations.

2.5.8 Exit Questionnaire and Debrief Session

As noted in Section 2.4.4, the exit questionnaire was administered on the final day of the study to those participants who took part in game-based training (i.e., the game-based section and OPFOR). The exit questionnaire was used to collect quantitative and qualitative data from participants regarding their experiences with game-based training and their opinions about its effectiveness. A debrief session was conducted immediately after the participants completed the exit questionnaire to discuss their responses in detail. At the completion of the session, all participants and assessors were given an overview of the preliminary findings from the study.

2.6 Data Analysis

All rating data were entered into a statistical database for analysis. The data were examined for any extreme values and deviations from normality. Descriptive statistics, including frequencies, means and standard deviations, were then calculated for the variables of interest (e.g., length of military service, confidence ratings, computer game experience, pre- and post-training scores). Similar analyses were conducted on each section's within-training data.

Due to non-normality of the data distributions non-parametric tests were used for all statistical analyses, unless otherwise stated. An alpha level of .05 was used for all significance testing. Exact probability (p) values are reported for all statistical tests except where $p < .001$. In figures, error bars represent the Standard Error of the Mean (SEM). Effect sizes, where reported, were calculated using Hopkins' (2002) criteria.

The results of the analyses for the demographic data, training data, and pre- and post-training assessments are outlined in the following sections. In addition to statistical significance, we also consider the practical significance of values. This involves consideration of the impact and usefulness of results of statistical testing. As noted by Kirk (2003), a difference could be statistically significant, but too small to warrant interest. A practically significant result is one where the result is both statistically significant, and of sufficient magnitude to warrant interest. In this study, we define a mean difference of 0.5 units or greater to be practically significant, which is halfway between successive points on the rating scale.

3. Results

3.1 Demographic Data

In this section, selected demographic data from the background questionnaire are presented; the complete set of demographic data are summarised in Appendix A.1.1. While some demographic data were collected from the OPFOR participants, it is not included in this report, as the population of interest was the two sections.

3.1.1 Length of Military Service

The participants' average length of military service – excluding the two Section Commanders - was 3.75 months (SD = 1.75) for the game-based section and 3.38 months (SD = 0.52) for the field-based section. A Mann-Whitney test showed that the difference in military service was not statistically significant, $U = 29.5$, $n_1 = n_2 = 8$, $p = .80$.

3.1.2 Computer Game Experience

Figure 4 shows the frequency distribution of computer game playing for each section. As shown, the profiles of game-playing frequency for each section were similar. On average, the participants in both sections played computer games on a weekly basis. Only one participant, in the field-based section, reported having never played computer games. The majority of participants reported that they played first person shooter (FPS) games, although the frequency with which they played these games was not assessed.

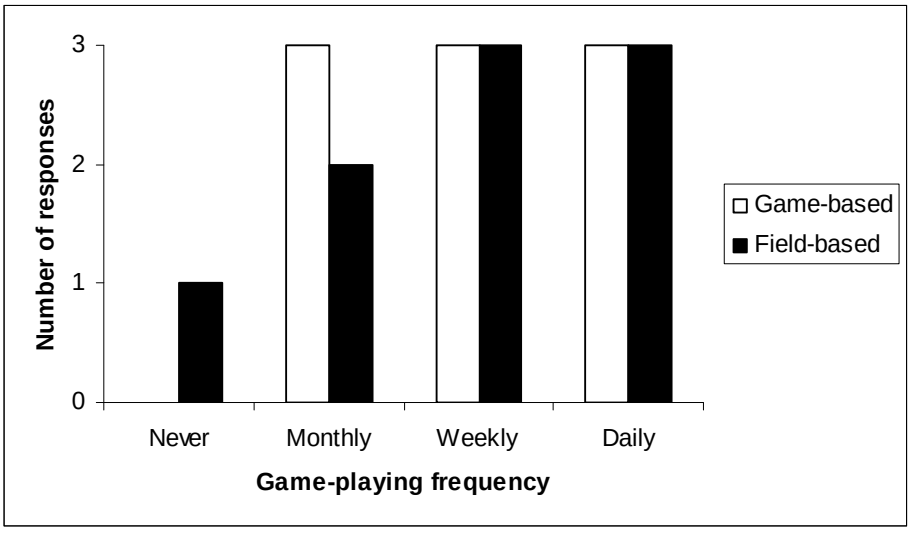

Figure 4: Frequency of computer game-playing for each section

3.1.3 Confidence in Conducting Section Attack Procedures

Figure 5 shows the self-reported levels of confidence in conducting section attack procedures for each section prior to training. As shown, the field-based section had a higher number of participants rating themselves as 'somewhat' or 'moderately' confident in section attack procedures. The two participants who reported they were 'very' confident were the Section Commanders, who each had 10 years of service in the Australian Army. To examine whether confidence levels differed across sections, each descriptor was assigned a numeric value (*not at all = 0, a little = 1, somewhat = 2, moderately = 3, and very = 4*), and the mean confidence rating of each section was calculated. A Mann-Whitney test showed that the difference between the mean confidence ratings of the two sections was not significant, $U = 23.5$, $n_1 = n_2 = 8$, $p = .38$.

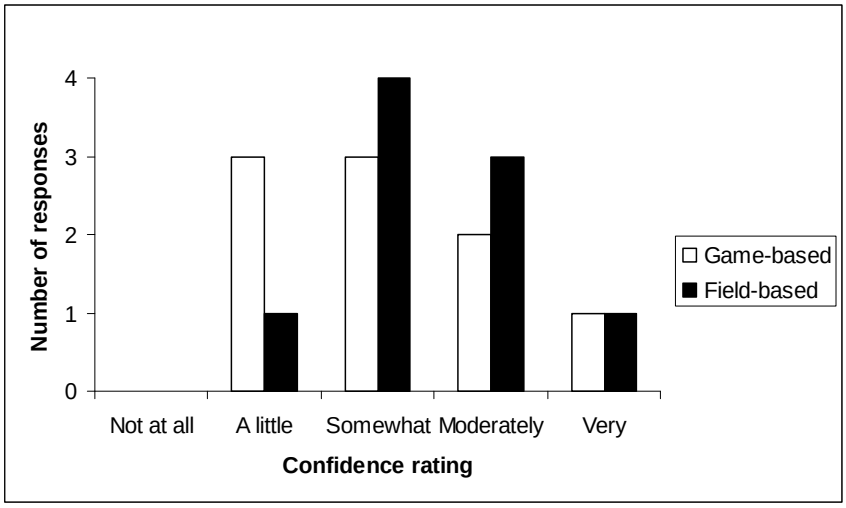

Figure 5: Self-reported confidence in conducting section attack procedures

In summary, the demographic data show that the two sections were similar on key variables such as length of service, computer game playing frequency, and confidence in conducting section attack procedures.

3.2 Pre-Training Data

The mean scores for each section on the pre-training assessment are shown in Figure 6. It is clear from the figure that the two sections were similar in their pre-training levels of performance. The figure also shows that both groups' performance level was quite low (approx. 1.5 units, max=5) equating to a level of performance between very poor (1) and poor (2) on the rating scale. A Mann-Whitney test confirmed that the difference in mean scores was not statistically significant, $U = 9647$, $n_1 = n_2 = 140$, $p = .82$.

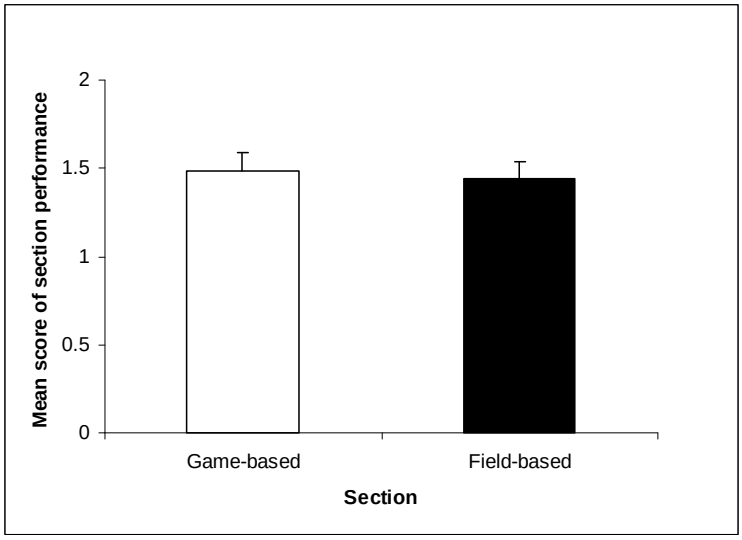

Figure 6: Mean scores on pre-training assessment for each section

The pre-training scores for each section were further broken down into the five behaviour categories on the assessment checklist (i.e., communication, leadership, coordination, supporting behaviour, and general points; Table 2). These data are presented in Figure 7. The pattern of scores for both sections is similar: coordination and supporting behaviour had the lowest scores, while the scores for communication, general points, and leadership were higher. The high score for leadership might be explained by the presence of the experienced Section Commanders. Given that the pattern is similar for both sections, it is possible that the data simply reflect the fact that novice trainees have higher levels of skill (albeit still rudimentary) in some areas than others. This possibility is not explored further in this report, although one of the assessors indicated that this was a plausible explanation.

Figure 7: Mean scores for each behaviour on pre-training assessment for both sections

A series of Mann-Whitney tests were conducted to see if there were any differences across the behaviour categories between the two sections. The results, summarised in Table 5, showed that there were no significant differences for any of the behaviours.

Table 5: Summary of Mann-Whitney testing of behaviour differences between sections

Behaviour	Mann-Whitney U	N_1, N_2	p
Communication	214	24	.11
Leadership	235	24	.26
Coordination	477	32	.62
Supporting Behaviour	282	24	.89
General Points	575	36	.39

The data were not analysed at the sub-section level (i.e., command group, gun group, and assault group) for two reasons. Firstly, as each group was assessed by a different assessor, it is possible that any differences between groups would reflect differences in assessment style rather than genuine performance differences. Secondly, as section attack is normally trained and assessed as a whole (i.e., as a section), it was not considered appropriate to analyse the performance of specific groups.

3.2.1 Competency-Based Assessment

Much of the Australian Army's training employs the principle of competency-based assessment, where performance is rated as either 'competent' or 'not yet competent' at the end of training. The baseline performance levels of the two sections were analysed using a competency-based approach, in order to determine if the patterns of results were similar. The data were reanalysed by assigning a score of 0 for any rating less than 3, and a score of 1 for any rating of 3 or more. In other words, a rating of 1 corresponds to a particular behaviour being performed to a satisfactory (i.e., competent) standard or better. The results are summarised in Table 4, containing both the mean score, and the percentage of ratings corresponding to a competent performance. As the competency-based assessment was a dichotomous score, either 0 or 1, the mean score multiplied by 100 equals the percentage of ratings corresponding to a competent performance. A series of Chi-square tests on these scores showed no significant difference between sections for each behaviour, as found above for the 0 - 5 rating scale.

Table 6: Mean pre-training scores for each behaviour category using a competency-based rating scale. The numbers in parentheses represent the percentage of assessment ratings corresponding to a competent performance.

Behaviour Category	**Game-based section**	**Field-based section**
Communication	0.25 (25.0)	0.42 (41.7)
Leadership	0.46 (45.8)	0.38 (37.5)
Coordination	0.06 (6.3)	0.03 (3.1)
Supporting Behaviour	0 (0)	0.04 (4.2)
General Points	0.44 (44.4)	0.33 (33.3)
Total	0.25 (25.0)	0.24 (23.6)

3.3 Training Data

3.3.1 Game-Based Training

The mean scores for the six training missions are shown in Figure 8. It is clear that there is a steady improvement in performance across the missions. Analysis using Friedman's test showed this change in performance was statistically significant, X^2 (5) = 35.01, $p < .001$. In addition, a Wilcoxon Signed-rank test showed that there is a significant difference between the mean scores for Missions 1 and Missions 6, $W = 1$, $Z = -2.39$, $p = 0.05$, indicating that the section's performance had improved by the end of training. Since performance increased by more than one rating point from the first to last mission (by 1.3 units) the results were also practically significant by our definition.

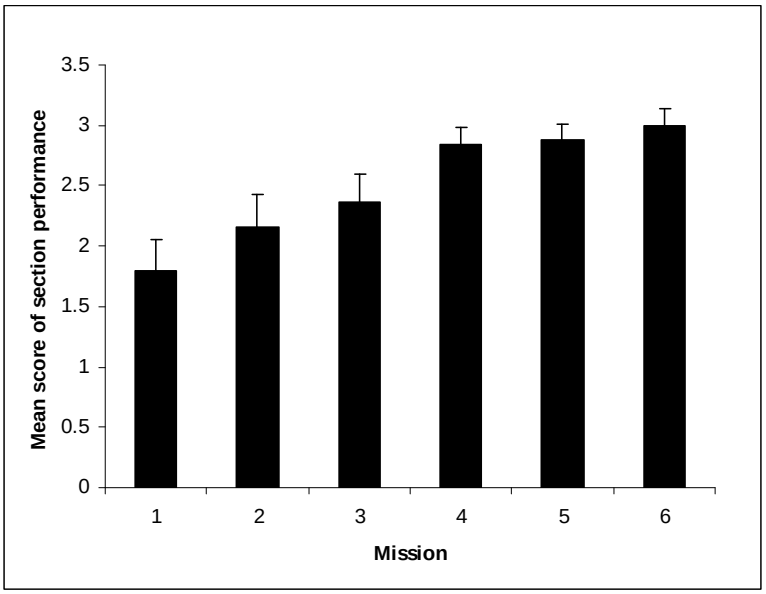

Figure 8: Mean scores of section performance for each mission during game-based training

To explore these data in more detail, the scores for each mission were broken down into mean scores for each behaviour category. These data are shown in Table 6. It can be seen that, while the scores for all behaviours increased across missions, the greatest change in scores was observed for behaviours in the general points category. The smallest change was observed for leadership behaviours, which is most likely due to the experience level of the Section Commander. These data were not subject to statistical analysis due to the small number of data points in each behaviour category.

Anecdotal observations from the assessor and researchers noted that while the section's performance appeared to be improving during training, there were still some problems with their execution of the section attack at the end of training. This included the Section Commander having difficulty keeping the section in formation and appropriately spaced, which lead to poor situation awareness at various times during the mission. This appeared to be a combination of participants experiencing difficulty controlling their avatars in the virtual environment, and lag issues resulting in computers refreshing at different rates.

Table 7: Mean scores for each behaviour across missions during game-based training

	Mission						Change from Mission 1-6
Behaviour	**1**	**2**	**3**	**4**	**5**	**6**	
Communication	1.5	3.0	3.0	2.8	3.0	3.0	+1.5
Leadership	3.4	3.6	3.4	3.4	3.2	3.6	+0.2
Coordination	2.3	2.3	2.6	2.9	2.9	3.0	+0.7
Supporting behaviour	1.3	1.5	2.0	2.3	2.3	2.3	+1.0
General points	0.2	0.4	0.8	2.8	3.0	3.0	+2.8
Average	**1.8**	**2.2**	**2.4**	**2.8**	**2.9**	**3.0**	**+1.2**

3.3.2 AAR Data

Preliminary examination of the AAR data revealed that there were only a small number of data points for each statistic (hit ratio, etc.). For instance, in four out of the six missions, the number of section members that fired rounds was four or fewer. The data sets were too small for statistical analysis, and as a consequence, the AAR data are not presented here in full[12]. Instead, only the hit ratio data are shown, to give a brief picture of the utility of the AAR data. These data are presented in Figure 9. For comparison, the assessor's ratings (from Figure 8) are shown together with the AAR data.

As shown in Figure 9, it is evident that the BLUEFOR and OPFOR hit ratios vary considerably across missions. The high BLUEFOR hit ratio in Mission 6 is due to one OPFOR being killed by one participant with a 100% hit ratio (1 round fired for 1 enemy wounded), and the remaining two OPFOR killed by a participant with a 33% hit ratio (6 rounds fired for 2 enemy killed). Averaging these values gives a hit ratio of 67%. However, this does not reflect the fact that other members of the section fired numerous rounds during this mission without killing or wounding any OPFOR. The low hit ratio in Mission 4 is due to an extremely high number of rounds fired by BLUEFOR (over 1000, an order of magnitude greater than in any other mission). This mission was of similar duration to the others, so the increased number of rounds fired appears to stem from different behaviours rather than a longer-lasting mission. This may reflect the use of excessive suppressive fire by section members. From these data, there are no obvious trends apparent, nor does there appear to be any correlation between the assessor's ratings and the AAR data. The utility of the AAR data are discussed further in Section 4.2.1.

[12] Refer to Whitney, Temby, and Stephens (2010) for a more detailed discussion of the AAR data.

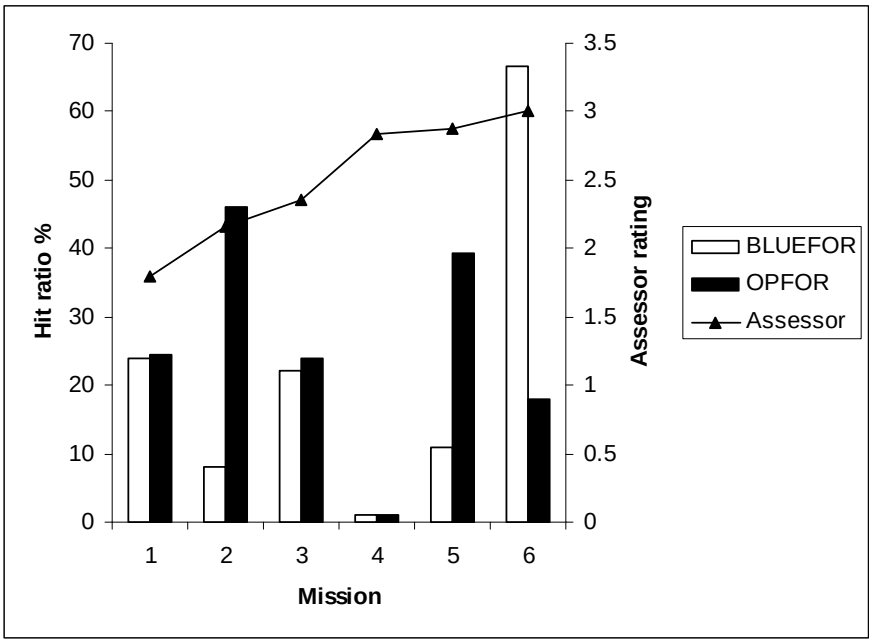

Figure 9: Hit ratio data (as percentage) for BLUEFOR and OPFOR during game-based training. The assessor's ratings are plotted on the secondary axis.

3.3.3 Field-Based Training

The mean score for the three section attack missions conducted during field-based training was 2.48. The assessor confirmed that training was conducted in a competent manner, although the performance level of the field-based section was still below a satisfactory level at the completion of training; this issue is discussed further in Section 4.1.

3.4 Post-Training Data

The mean scores for both sections from the pre-training and post-training assessments are shown in Figure 10. It is clear from the data that the field-based section shows substantial improvement from pre-and post-training assessment. However, there is little difference between the pre- and post-training scores for the game-based section. Given the large sample size (140 data points per condition), an ANOVA was used for analysis, as it is robust against violations of normality with sample sizes over 100 (StatSoft, 2010). T-tests were used for post-hoc analyses for the same reasons. A 2 x 2 mixed ANOVA showed a significant effect for test time (pre- vs. post-), F (1, 278) = 47.59, $p < .001$, and significant interaction for test time and section, F (1, 278) = 35.71, $p < .001$. Post-hoc testing using an Independent Samples T-test confirmed that the field-based section's performance was significantly greater than the game-based section's performance on the post-training assessment with a moderate effect size, t (278) = 7.84, $p < .001$, $d = 0.69$. A Paired Samples T-test confirmed that the field-based training section's performance increased significantly following training with a moderate effect size, t (139) = 9.52, $p < .001$, $d = 0.83$. No other comparisons were significant.

As noted previously (Section 3.3.1), current limitations with VBS2 restrict the participants' ability to execute certain skills required for section attack. To investigate whether this factor discriminated against the game-based section, additional analyses were conducted. The pre- and post-training scores for both sections were recalculated after removing the ten items identified as being generally too difficult to be assessed or conducted in the virtual environment (see Table 3). The post-training score for the game-based section did not improve as a result of removing these items; in fact, the post-training score decreased slightly, from 1.55 to 1.51. Given that recalculating the score produced such a negligible change in the mean score, these data were not subject to further statistical tests.

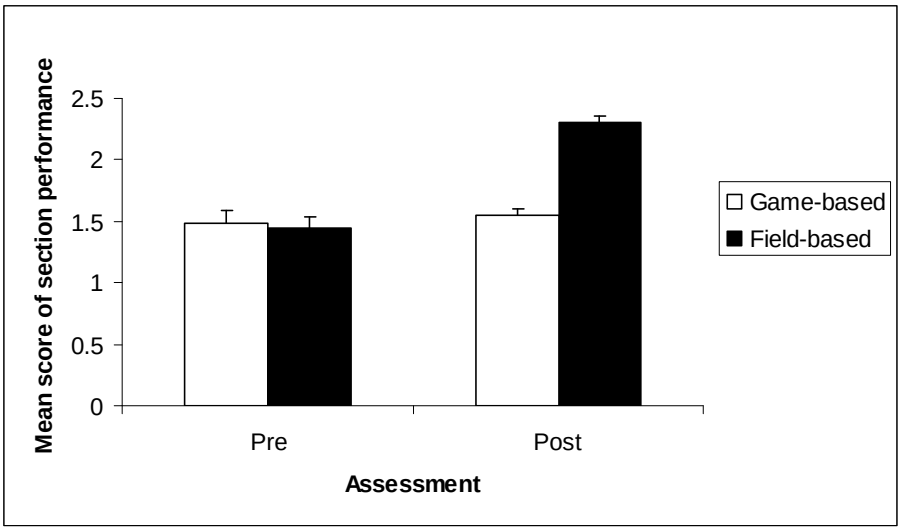

Figure 10: Mean scores for pre- and post-training assessments for both sections

To obtain greater insight into the changes occurring from pre- to post-training, the frequency distribution of the change in rating scores was plotted for each section[13]; the results are shown in Figure 11. For the game-based section, the data are symmetrically distributed about zero; that is, for many items, there was no change in score. Nevertheless, there were still many instances where the score increased or decreased, in accordance with the section showing inconsistency in their performance, as might be expected with inexperienced soldiers. Overall, however, there was no net change in the score. In contrast, the distribution of ratings for the field-based section is skewed to the right; there are far more instances of positive rating changes than negative, consistent with the observed overall increase in performance for this section.

[13] This was calculated by subtracting the pre-test score from the post-test score for all items for each of the three post-training section attacks

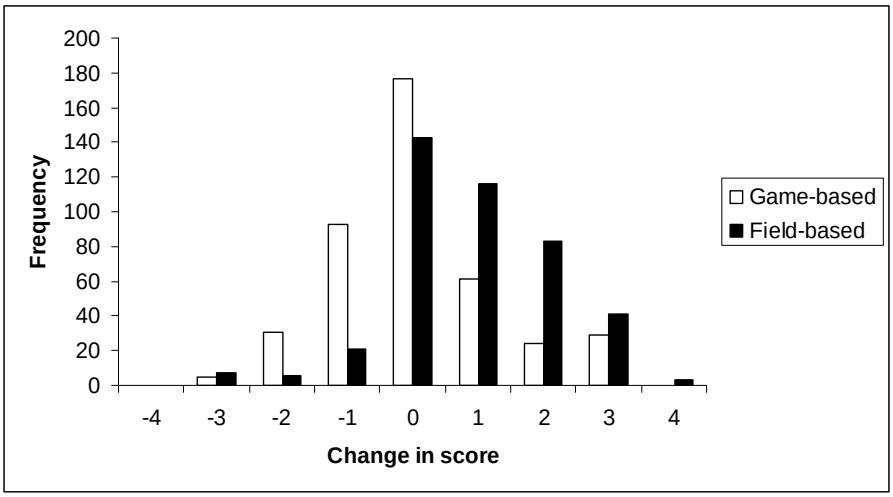

Figure 11: Frequency distributions of changes in score from pre- to post-training for each section

Because there were differences in the pre-training data for the five behaviour categories, the post-training data were also analysed by behaviour (Figure 12). Wilcoxon Signed Ranks were conducted on the pre- and post-training scores for each section. Results of the tests, summarised in Table 8, indicate that the game-based section experienced a significant *decrease* in communication scores following training. The field-based training section experienced significant increases in all behaviours except Communication.

Table 8: Results of Wilcoxon Signed Ranks tests for pre- and post-training scores

	Game-based section		**Field-based section**	
	Z score	***p* value**	**Z score**	***p* value**
Communication	-3.84	.701	-1.17	.24
General Points	-2.29	.022	-3.03	.002
Supporting Behaviour	-1.73	.084	-3.95	.001
Leadership	-.43	.664	-3.75	< .001
Coordination	-1.88	.059	-4.93	< .001

Taking a change in performance of 0.5 units or more as practically significant (see Section 2.6), the positive performance changes observed for the field-based section are practically significant, except for communication, which approaches this criterion. In contrast, none of the performance changes for the game-based section were practically significant.

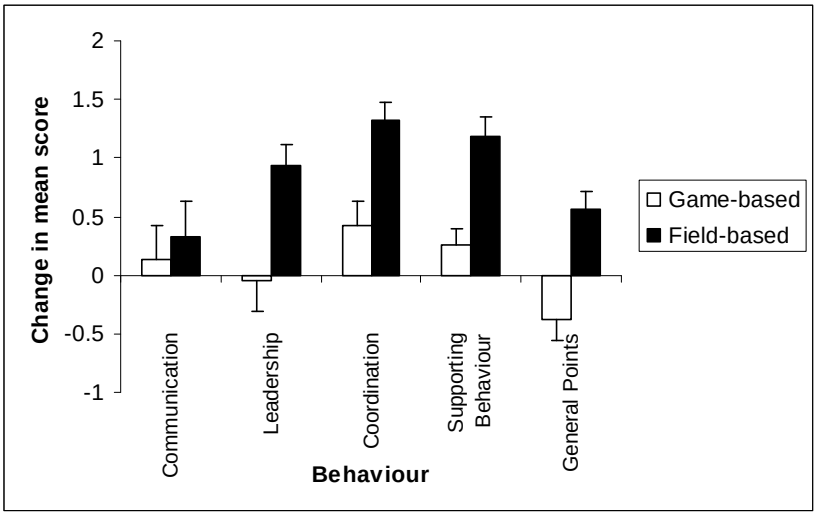

Figure 12: Difference between mean pre- and post-training scores for each behaviour category

3.4.1 Impact of Non-Blind Assessors

As discussed in Section 2.2 and Footnote 7, two of the four assessors were not blind to the type of training undertaken by the two sections. In order to assess whether or not this had any impact on the post-training assessment, the pre- and post-training scores were analysed according to whether the assessors were blind or not (Figure 13). Some variability between assessor ratings is expected, given that each assessor was positioned in a different field location during the assessments (see Section 2.5.4). The differences between mean scores for blind and non-blind assessors were not statistically significant ($p > 0.05$) except for the game-based post-training scores. In this case, the non-blind assessor mean scores were 0.4 points higher than the blind assessor scores ($p < 0.001$); however the difference is not practically significant. This suggests that the assessment conducted by the non-blind assessors was not biased by their knowledge of the type of training each section received..

As an additional examination of the potential impact of non-blind assessors, Cronbach's alpha values were calculated for each section's pre- and post- testing scores. These values and summarised in Table 9. Using the criteria suggested by George and Mallery (2003), internal consistency is good for the game-based training section's post-training score, and acceptable for the remainder. Consequently, we are confident that the overall findings are reliable despite not all assessors being blind to the training methods

Table 9: Cronbach's alpha values for each section's pre- and post-training scores

	Game-based	Field-based
Pre	.733	.753
Post	.803	.722

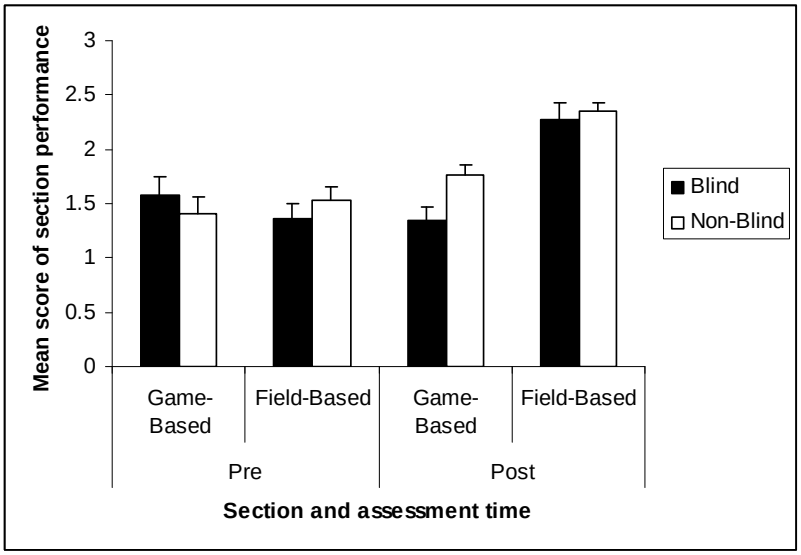

Figure 13: Mean pre- and post-training scores for both sections, comparison of blind and non-blind assessors

3.4.2 Competency-Based Assessment

As with the pre-training data, the post-training data were reanalysed by assigning a score of 0 for any rating less than 3, and a score of 1 for any rating of 3 or more. Consequently, a rating of 1 corresponds to a particular behaviour being performed to a satisfactory (i.e., competent) standard or better. The results are summarised in Table 7 and do not offer any further insights beyond those obtained using the 0 – 5 rating scale[14]. Overall, only the field-based section showed any significant improvement following training, with a satisfactory level of performance achieved on 55.5% of all criteria.

Table 10: Mean pre- and post-training scores for each behaviour category using a competency-based rating scale. The numbers in parentheses represent the percentage of assessment ratings corresponding to a competent performance.

Behaviour Category	Game-based section		Field-based section	
	Pre	Post	Pre	Post
Communication	0.25 (25.0)	0.21 (20.8)	0.42 (41.7)	0.51 (51.4)
Leadership	0.46 (45.8)	0.43 (43.1)	0.38 (37.5)	0.71 (70.8)
Coordination	0.06 (6.3)	0.19 (18.8)	0.03 (3.1)	0.47 (46.9)
Supporting Behaviour	0 (0)	0.07 (6.9)	0.04 (4.2)	0.51 (51.4)
General Points	0.44 (44.4)	0.24 (24.1)	0.33 (33.3)	0.58 (58.3)
Total	0.25 (25.0)	0.23 (22.6)	0.24 (23.6)	0.55 (55.5)

[14] The statistical significance of differences between pre- and post-scores for each behaviour were measured using a chi-squared test based on the frequency of 0 and 1 ratings. The results were essentially the same to those obtained using the 0 – 5 rating scale.

3.4.3 Knowledge-Based Assessment

The results from the pre- and post-training assessments (Section 3.4), which show no significant change in performance after training for the game-based section, are in contrast to the results obtained during the game-based training sessions (Section 3.3.1), which clearly showed that performance improved during training (i.e., learning was occurring). In order to explore this further, the data were analysed to determine if the game-based training section's level of knowledge improved during training. This was done by using a rating scale in which scores greater than 0 were assigned a value of 1. Such a scale simply measures whether a task or behaviour was attempted, regardless of how well it was carried out and hence is potentially useful in detecting knowledge-based learning, as opposed to measuring an increase in skill level as a result of training.

In other words, a rating of 1 was given when a particular behaviour is demonstrated. The results of this analysis are summarised in Appendix C. In this instance, statistical significance was determined using a Chi-squared test based on the frequency of 0 and 1 ratings (see Table 15 in Appendix C). Both sections were equivalent on the pre-test (as with the other rating scales used in this report), but in this case, both sections showed significant post-training improvement. When analysed by behaviour (Figure 14) significant performance improvements were observed for the game-based section for all behaviours (in contrast to results observed when using the 0 – 5 rating scale). The magnitude of these changes is worthy of interest (> 10 %) for all behaviours except for general points. Overall, the game-based section showed an improvement in their knowledge levels (based on the frequency of ratings of 1 as a percentage of the total) from 71.4% to 86.4% after training, which is large enough to be considered noteworthy[15].

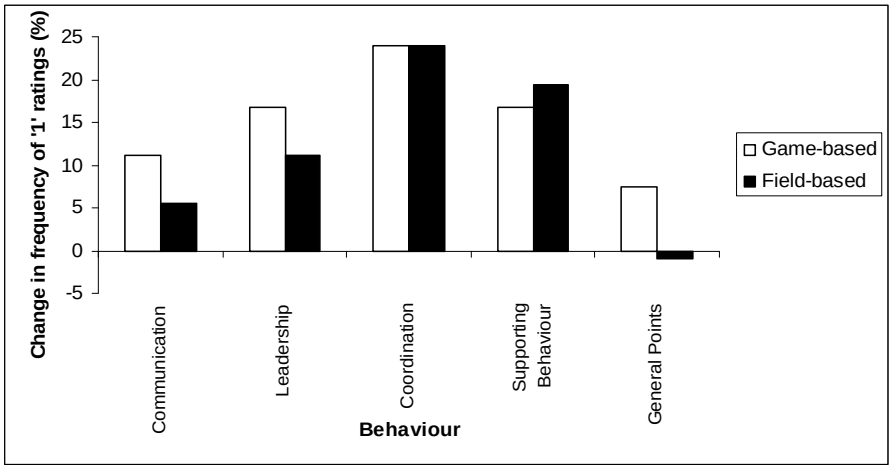

Figure 14: Difference between pre- and post-training performance for each behaviour using a knowledge-based rating scale.

[15] The changes for the field-based section were similar. However, there is little value in examining the field-based section data using this rating scale, because it does not discriminate between any score above zero so the real differences between the two sections are not apparent. In other words, we already know that the field-based training section's performance improved.

From this analysis, there is some evidence of knowledge transfer of section attack procedures for the game-based section. In order to investigate the extent to which this could actually be attributed to the use of VBS2, the data were analysed after removing the ten items identified as being too difficult to be adequately performed in VBS2 (see Table 3, Section 2.4.2). The results of this analysis are presented in Appendix C and in Figure 15 below. It is apparent that removing these items actually results in a smaller performance increase for the game-based section; overall, the performance increase for the game-based section has reduced from 15.0% to 8.3%. The significance of this finding is discussed in Section 4.2.

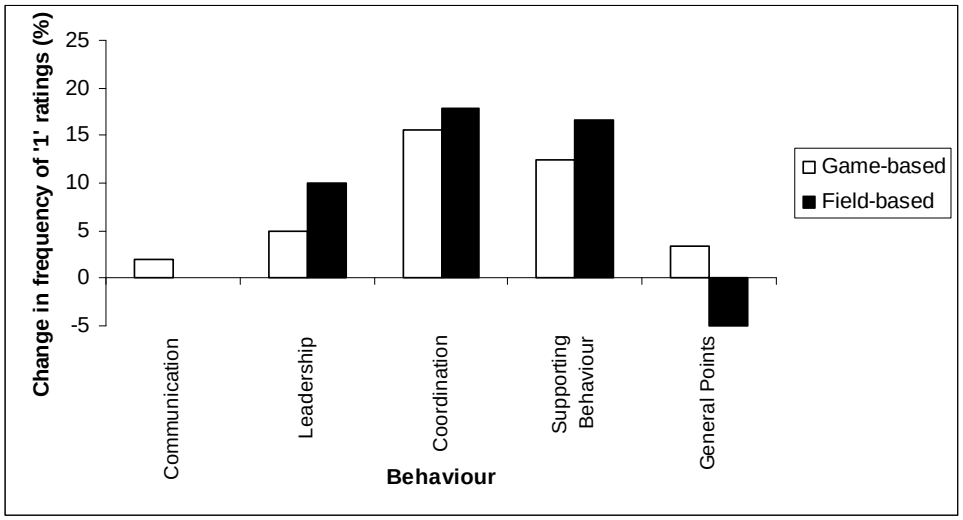

Figure 15. Difference between pre- and post-training performance for each behaviour using a knowledge-based rating scale after removing 10 items from assessment criteria.

3.4.4 Exit Questionnaire Data

The exit questionnaire data are summarised in Table 11. These data are based on the responses of the game-based section (n=9) but not the OPFOR. The exit questionnaire data show that the majority of participants rated navigating in the virtual environment and communicating with other section members to be easy. Approximately half of the participants reported that fire and movement was somewhat difficult in VBS2, while the other half rated it somewhat easy. In addition, the majority of participants reported that it was difficult to maintain situational awareness of their team member's positions and to identify the OPFOR avatars in VBS2. Over two-thirds of the participants indicated it was somewhat difficult to achieve the mission objectives; the remainder reported it was somewhat easy. Overall, these data provided some indication of the difficulties that participants experienced when using VBS2 to conduct section attack missions; these difficulties are discussed further in Section 4.2.

Table 11: Summary of exit questionnaire data shown as frequency of responses for each activity

Activity	Very Easy	Somewhat Easy	Somewhat Difficult	Very Difficult
Navigate	2	5	2	0
Communicate	2	6	1	0
Fire and Movement	0	4	5	0
Situation Awareness	0	1	6	2
Identify OPFOR	0	0	5	4
Achieve Mission Objectives	0	2	7	0

All participants involved in game-based training indicated that they believed they received enough training time on the game, and that the game was generally easy to learn. With regard to the perceived effectiveness of game-based training, five participants rated it as 'moderately effective', two rated it as 'somewhat effective' and two rated it as 'not at all' effective. Therefore, most participants rated the training as being *at least somewhat effective* in preparing them for the section attack activities in the field; however the post-training assessor ratings clearly showed that the skills of interest did not transfer to the field.

During the debrief session, when the participants were given the opportunity to discuss their survey responses, it was clear they did not believe that using VBS2 for section attack training was very effective in preparing them for the field activities. Overall, the commander of the game-based section commented that he did not believe the game provided any training benefit, with the exception of perhaps helping to teach section formations.

The most common things that participants liked about VBS2 were: the 2D view for visualising section formations, the 'realistic' representation of ADF equipment in the virtual environment, and the different mission scenarios that could be created. In contrast, the participants disliked the limited field of view, the inability to isolate the direction of enemy fire, and the on-screen resolution of the terrain and objects in the virtual environment.

In terms of modifications that participants would make to improve the game as a training tool, a number of suggestions were noted. The most common suggestion was the need to improve target indication; participants had great difficulty in identifying the direction that enemy fire was coming from. This had a significant impact on the ability of the section to react appropriately in these situations. It was also suggested that the quality of the graphics and animation needed to be improved; it is possible that the participants were rating VBS2 unfavourably compared to other FPS games they had played. This is consistent with comments reported by participants in other studies (e.g., van der Hulst et al, 2008).

4. Discussion

The aim of this study was to examine the effectiveness of a desktop computer game to train dismounted soldiers in infantry section attack procedures, including investigating whether there was any measurable training transfer to the field. In the following sections we discuss the findings in relation to each of the study objectives.

4.1 Comparison of Game-Based and Field-Based Training Outcomes

The first objective of the study was to compare training outcomes for game-based and field-based training. This was achieved by examining assessor ratings of each section's performance during and after training. The data confirmed that the field-based section's performance increased significantly from baseline levels following training. On the basis of their post-training score, the field-based training section achieved a score between 'Poor' and 'Satisfactory', and one of the assessors indicated that with an additional day of training, their performance would probably have reached a satisfactory level. This finding provides evidence that indicates that the current method of training section attack procedures used by Australian Army is effective, as would be expected.

In contrast, while the game-based section showed performance improvement during training, this did not translate to improved performance in the field on post-training measures. The game-based section's performance on the post-training assessment equated to a rating between 'Very Poor' and 'Poor'. This finding is in contrast to previous studies of game-based training, such as those reviewed in the Introduction (Section 1.2). In general, these studies have reported similar post-training levels of performance for groups undertaking game-based or field-based training and taken this as evidence for the effectiveness of game-based training. However, as highlighted in the Introduction (Section 1.2), and discussed further in Section 4.3.1, such findings are more likely the result of methodological limitations rather than genuine effects.

4.2 Evaluation of the Effectiveness of VBS2 for Training Section Attack

The second objective of the study was to evaluate the effectiveness of the VBS2 game for training novice infantry personnel in section attack procedures. This objective was investigated by comparing post-training scores for the two sections (based on assessor ratings), and examining participants' perceptions of their training experiences with VBS2. From the game-based training data, the increase in performance ratings provided evidence that learning was occurring. However, the post-training data show that despite this, whatever skills were learnt during training did not transfer from the virtual environment to the field. In fact the only evidence that there was any benefit from the game-based training sessions was found when a knowledge based assessment of the data was conducted (see Section3.4.3). This showed that some of the knowledge learned during the game-based training did transfer to the field, although the improvement was modest (15%). Indeed, when the data were modified to exclude those items difficult to assess or conduct in a virtual environment, the effect was even smaller (8%). This suggests that the effect was probably due to the fact that participants' knowledge of section attack was improved during training through repetition and regular feedback from the Section Commander, rather than being due to any specific training benefit afforded by the game. Overall, the assessor ratings suggest that VBS2 has limited effectiveness for training novice infantry teams in section attack procedures, especially when contrasted with the performance improvements seen for the field-based section.

The exit questionnaire data showed that the majority of participants rated the game-based training as being at least 'somewhat effective' as preparation for the subsequent section attacks conducted in the field. However, when given the opportunity to elaborate on their

ratings in the debrief session, none of the participants indicated that training with VBS2 provided any benefit in preparing them for the field activities, consistent with the assessor ratings on the post-training assessment. This discrepancy between participants' initial ratings of training effectiveness and their subsequent comments in the debrief highlights the risk of relying solely on participant ratings and the importance of more objective data (i.e. assessor ratings) when evaluating training effectiveness.

The exit questionnaire and debrief data, together with observations by the study team and assessor during the game-based training missions provided specific information on aspects of VBS2 that limit its effectiveness for collective/team training. The major aspect is the difference between the physical skills required to conduct section attack in the field, compared with the virtual environment. In the field, a section attack involves running, going to ground, crawling and finding cover while maintaining awareness of other team members, as part of coordinated fire and movement. These are all highly physical skills that also involve cognitive processes; this is consistent with analyses of infantry tasks conducted by other researchers (Nolan & Jones, 2005; Tack & Angel, 2005), and the authors' analysis of a section attack. The tempo of the section attack is such that both the cognitive and physical skills need to be executed almost simultaneously. Consequently, if we accept this reasoning, training the two sets of skills separately may not be effective. We do not speculate further on this issue; it is an area for future research to investigate in more detail.

Other specific limitations relating to VBS2 are listed below.

- Field of view: the limited field of view in VBS2 had an impact on participants' peripheral vision and spatial awareness of their team members' location; in particular, this made it difficult for the Section Commander to control the section. The effect of a limited field of view on soldier performance in virtual environments has been highlighted by previous researchers (e.g., Lewis, 2005; Morrison et al., 2005; van der Hulst et al., 2008). Given the critical role of the Section Commander, this limitation was a major impediment in using VBS2 to conduct and train section attack procedures.

- Target indication and detection: the participants had problems with target detection and indication, which made it difficult for them to isolate enemy positions in the virtual environment. This problem was related to the lack of directionality in the audio cues. This could potentially be addressed by incorporating 360° sound effects into the game, but its effectiveness would depend on the extent to which a computer's speaker system supported multidirectional sound effects.

- Cover and concealment: participants experienced difficulties in maintaining adequate cover and concealment from enemy fire during the missions. While the terrain used in the missions provided areas for cover, participants were often unable to determine the extent to which their avatar was adequately protected from enemy detection or fire by the terrain features, especially if they were unable to identify the enemy's precise location.

- Weapon sights: there was an over-reliance by some participants on using weapon sights to scan the virtual environment. While the magnification in the sights made it easier to locate enemy positions in the distance, the use of the weapon sights further reduced the field of

view, which often led to those participants becoming separated from the other members in the section when advancing toward the enemy.

Overall, while VBS2 was able to be used to conduct section attack procedures, the above factors made it difficult for the section to carry out the missions successfully even when the actions of the OPFOR were heavily constrained to ensure they supported the training objectives. The findings suggest that VBS2 in its current configuration has limited effectiveness as a tool for training novice infantry teams in section attack procedures. The findings do not imply that VBS2 has no training utility at all; rather it means that alternative uses for VBS2 should be investigated and evaluated. This issue is discussed further in Section 4.6.

4.2.1 AAR Data

As discussed in Section 3.3.2, no statistical analyses were undertaken on the AAR data. The AAR data gave a different picture of the sections' performance during the game-based training, when compared with the SME data. Specifically, the SME data indicated that the section's performance improved across the missions, whereas the AAR data showed that performance was variable across the missions with no clear trends. Consequently, there was no clear relationship between the two sets of data. There are several plausible reasons for this finding.

Firstly, it is apparent that the hit-ratio is not a good measure of the section's performance due to the use of suppressive fire in several of the missions, as noted in section 3.3.2. Secondly, the AAR data provides statistics regarding engagement outcomes at the individual and team level, whereas the SME data provides information regarding all team processes undertaken by the section. The lack of any clear relationship between the two sets of data is likely due to these differences. This view is discussed in more detail in Whitney, Temby, and Stephens (2010), and is consistent with the view of Hussain and Feurzeig (2008) who suggest that current AAR tools in serious games are insufficient for accurate assessments of individual and team performance. The summary statistics generated were not useful measures of the section's performance. However, the AAR tool had some use in reinforcing specific teaching points, such as the importance of maintaining adequate spacing between section members.

Overall, these findings highlight the need to develop better objective measures of task and team performance within the AAR capabilities of computer games. Details of how this might be achieved are described further in Whitney et al. (2010).

4.3 Methodology for Evaluating Game-Based Training

The third objective of the study was to evaluate a methodology for assessing game-based training. This objective was included in the study in light of methodological shortcomings in previous studies in this area. This objective was addressed in two ways: firstly by incorporating control measures into the research design and considering their impact on the study outcomes relative to previous studies, and secondly by examining how well the methodology met the criteria for a good experiment outlined in The Technical Cooperation Program (TTCP) Guide for Understanding and Implementing Defense Experimentation

(GUIDEx; TTCP, 2006). We now evaluate the methodology employed in the current study in the following sections using these two approaches.

4.3.1 Comparison with Previous Studies

As highlighted in the Introduction (Section 1.2), previous studies of game-based training have been hampered by methodological shortcomings that limit the strength of conclusions that can be drawn from them. Specifically, these shortcomings include:

- Use of experienced military personnel who are unlikely to benefit significantly from game-based training (e.g. Wiederhold, 2005)
- No pre-training measures of trainee performance (e.g. van der Hulst et al., 2008; Woodman, 2006)
- Insufficient time allocated to game-based training (e.g., Wiederhold, 2005)
- Non-blind assessment methods (e.g., Kneuper, 2006; Nolan & Jones, 2005)
- Use of rating scales and performance measures that are too simplistic and not sensitive enough to detecting changes in trainee performance (e.g., Kneuper, 2006; Woodman, 2006)
- Reliance solely on participant opinions (i.e., self-rating data) to evaluate game-based training outcomes (e.g., Nolan & Jones, 2005; van der Hulst et al., 2008)

The current study was successful in addressing each of the above shortcomings in the following ways:

- Use of inexperienced military personnel who were more likely to benefit from training
- Measuring pre-training performance, thereby allowing assessment of performance changes after training
- Adequate training time on game (i.e., eight hours over two days)
- Use of blind assessment methods to minimise the potential for assessment bias
- Use of rating scales and performance measures that were sensitive to changes in team performance levels
- Use of experienced assessors to rate participants' and team performance (rather than relying solely on self-report data from participants)
- Allocating participants to experimental groups using appropriate demographic variables (e.g., prior experience) to minimise any pre-existing differences between groups before undertaking training.

As a result we are confident that the findings (i.e. difference in training outcomes) are valid and due to differences in training methods and not extraneous variables. Overall, the methodology employed was arguably more rigorous than those used in previous game-based studies and we would recommend that a similar approach be adopted in future studies.

4.3.2 Comparison with TTCP GUIDEx

The TTCP GUIDEx (2006) is a document that has been published to assist researchers when designing and conducting studies in defence environments. The GUIDEx details 21 threats to

a good experiment under four categories: ability to use capability, ability to detect change, ability to isolate reasons for change, and ability to relate results to defence operations. In scientific terms, these are threats to internal and external validity. To allow another assessment of the methodology used in this study, we conducted an analysis of the 21 threats against the procedures we adopted in the study. The results of this analysis are summarised in Table 17 in Appendix D. Overall, this analysis showed that all of the relevant threats were mitigated with appropriate strategies. This analysis provided further confirmation that the methodology was rigorous and that the study had good levels of internal and external validity.

4.4 Implications of Findings

The final objective of the study was to identify the implications of the study findings in relation to using games for training. These implications are discussed below.

4.4.1 Cost-Benefit

In this study, the nominally cheaper method of training section attack was not found to be effective[16]. Consequently, game-based training with VBS2 would not be considered a cost-effective alternative to the current method of training dismounted soldiers in section attack (noting that VBS2 is not currently used by the Australian Army for this purpose). The question of whether VBS2 provides any training benefit to more experienced soldiers is beyond the scope of this study and possibly an area for future research to address.

4.4.2 Generalisability to Other Collective Skills

The findings from the current study are specific to using VBS2 to train dismounted infantry teams in section attack procedures. However, the skills required in section attack constitute approximately 50% of all skills required in the conduct of other infantry TTPs (e.g., patrolling, ambushing, obstacle crossing). This number was derived from an unpublished task analysis conducted by a Warrant Officer in the Australian Army with over 20 years of experience in infantry TTPs. All of these activities are highly physical and, in each case, the Section Commander plays a critical role and requires a high level of situation awareness. Given the results of this study, it is reasonable to conclude that VBS2 is not an ideal tool (in its current configuration) for training novices in dismounted infantry TTPs. More broadly, this raises the question as to whether any team tasks that require good peripheral vision and situation awareness for effective performance can be trained appropriately using VBS2 (and similar first-person perspective games) in its current configuration. This is potentially a useful area for future research to address.

[16] The initial resources required to acquire a game-based training capability (including infrastructure) can be significant. Once acquired, however, game-based training is generally less expensive than live training due to the cost of ammunition, transport of troops to field training areas etc.

4.4.3 Collective versus Individual Training

The findings from the study are specific to using VBS2 to train a specific team task with dismounted soldiers. Consequently, the implication is that no inferences should be made on the effectiveness of VBS2 as a tool for training individual skills or mounted soldiers (for example, to train individuals to drive military vehicles). Consequently, when discussing the effectiveness of game-based training, we believe it is important to distinguish between using games to train (a) individual skills and (b) collective/team skills. This distinction may be useful because certain simulations may be effective for individual training but not collective training (and vice versa). For example, while VBS2 appears to be ineffective for training section attack (which involves collective skills) it is widely regarded as an effective visualisation tool for terrain familiarisation (which is mainly an individual skill), where it is currently used by the Australian Army as part of pre-deployment training. It may also be beneficial to examine the effectiveness of VBS2 for training different types of skills, e.g. physical skills vs. cognitive skills. However, it is outside the scope of this report to address this in detail.

4.5 Limitations of Study

Despite the high level of scientific rigour achieved in this study, there were some limitations which should be addressed as part of future research.

Firstly, while the measures used in the study were useful in discriminating between the performance of the two sections, future research might consider including a specific test to assess participants' knowledge levels of the task being trained before and after training. As shown in this study, while there was little evidence of behavioural changes in the game-based section on post-training measures, there was some evidence of improved knowledge, albeit using indirect measures (i.e., by analysis, see Section 3.4.3). Consequently, to ensure that any knowledge acquired during training is captured, it may be worthwhile to include specific knowledge measures, such as written tests, in future studies (e.g., Ricci, Salas, & Cannon-Bowers, 1996).

Secondly, the findings from this study showed that while the field-based section performed better than the game-based section on post-training measures, the section's level of performance at the end of training was still below a level considered satisfactory. As noted previously, one of the assessors believed that with an additional day of training the field-based section would achieve a satisfactory level. While not a limitation of the study per se (indeed, individual differences in ability levels across cohorts of military trainees are natural and will impact on the time required for them to achieve competency), the issue here for future research is to factor in enough time to train to competency. Although training to competency was not a specific objective in this study (and therefore not critical to the outcomes) it is a point worthy of consideration in future studies where the standard or level of task performance achieved at the end of training is critical.

Finally, this study involved a comparison of two training types: game-based training and field-based training (i.e., traditional instruction). While efforts were made by the study team in the planning stages to include a third training condition, consisting of a mix of game-based

and field-based training, this was not possible due to manning shortages. In this study, the performance of a section undertaking game-based training was examined. While not a limitation per se, it is not possible to draw any strong conclusions from this study about the efficacy of blended training methods to teach section attack procedures. It is plausible that using a combination of field-based training and game-based training is an effective method of training because of the opportunities to reinforce knowledge and procedures in both environments. However, in the case of section attack, even if game-based training was found to contribute significantly to a blended training method, it is questionable whether it would be cost effective to implement such a solution (as discussed in Section 4.4.1). Field-based training for section attack is not overly costly; at the SOI, the section simply had to collect their weapons from the armoury and walk a short distance to the field training area. Conversely, there is a high initial cost involved in setting up a dedicated computer suite for game-based training, as well as ongoing overheads associated with technical support and training instructors in the use of the game. Given this issue, and the fact that game-based training alone was not found to be effective, this raises doubts over the value of repeating this particular study with a blended training condition added.

4.6 Future Research Directions

In summary, there are several areas for future research into game-based training;

- Investigate potential solutions to address the limitations with VBS2 identified in the current study, including limited field of view, difficulties in target identification and detection, maintaining adequate cover and concealment, and reducing the over-reliance on weapon sights. This may require the technical support of industry.
- Investigate the use of game-based training for different military tasks. For instance, recent research by Beal (2009) suggests that game-based training may be effective for training commanders in decision making, while other studies suggest that it may be useful for vehicle crew training (Roman & Brown, 2008; Whitney et al., 2012) and counter-IED training (Jarmasz et al., 2010).
- Examine the optimal mix of game-based and traditional training for specific military tasks. This would require a series of experiments in which the relative time spent on the different training methods and the sequence in which these methods were applied was varied.

5. Conclusion

This study has found two unique outcomes in the area of game-based training for small teams of dismounted soldiers. Firstly, it was found that traditional training produced significantly better outcomes than game-based training. Secondly, game-based training had no measurable benefit for training an infantry team task (i.e. section attack). Given the level of scientific rigour achieved, together with the effect sizes obtained, we are confident that our findings are reliable.

These outcomes are in contrast with previous studies with dismounted combatants, which have concluded that game-based training is effective. However, these studies were unable to

quantify the relative effectiveness of game-based training and traditional methods, as we have in the current study.

In conclusion, this study has found evidence of training transfer when traditional instruction methods are used to train dismounted soldiers in infantry procedures. In contrast, the results indicate that there is negligible transfer of these skills when using game-based training alone. Consequently, the findings suggest that the current method of training section attack procedures in the Australian Army is effective and that game-based training using VBS2 would not be effective for this task. Additional implications of the current findings have been discussed, and recommendations for future research into game-based training outlined; these recommendations include the need to determine the optimal mix of instructional methods for training other military skills, and addressing technological limitations of games in order to better support training requirements.

6. Acknowledgements

We thank the Commanding Officer of the SOI for his support to the study, the staff from Rifleman Wing who assisted in planning and conducting the study, and HQ Training Command - Army for their assistance in gaining access to the study participants. We are grateful to WO2 Paul Dabinet and WO2 John Williams for their help with the assessment criteria and training scenarios, and their assistance with data collection. We also thank Dr Christopher Shanahan from Air Operations Division for his assistance during the study. Finally, we thank all the participants for their enthusiasm and cooperation during the study.

7. References

Atherton, E., & Baxter, H. C. (2009). Positively gaming the system: A VBS2™ training case study. *Proceedings of the 2009 Interservice/Industry Training, Simulation, and Education Conference.* Orlando, Fl: NTSA.

Barlow, M., Morrison, P., Luck, M., & Dickie, A. (2004). Constructing the virtual section. *Proceedings of the 2004 SimTecT Conference*. Sydney: Simulation Industry Association of Australia.

Beal, S.A. (2005). *Using games for training dismounted light infantry leaders: Emergent questions and lessons learned (Research Report 1841).* Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Beal, S.A. (2007). *Assessment of two desk-top computer simulations used to train Tactical Decision Making (TDM) of small unit infantry leaders (Research Report 1869).* Arlington, VA: US Army Research Institute for the Behavioral and Social Sciences.

Beal, S.A. (2009). Exploring the use of a multiplayer game to train infantry company commanders. *Proceedings of the 2009 Interservice/Industry Training, Simulation, and Education Conference.* Orlando, Fl: NTSA

Beal, S. A., & Christ, R. E. (2004). *Training effectiveness evaluation of the Full Spectrum Command game (Technical Report 1140).* Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Belanich, J., Mullin, L.N., & Dressel, J. D. (2004). *Symposium on PC-based simulations and gaming for military training (Research Product 2005-01).* Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Bell, B., Jarmasz, J., & Nelson, I. (2011). Development of scenario-based pre-deployment counter-IED training. *Proceedings of the 2011 Interservice/Industry Training, Simulation & Education Conference*. Orlando, Fl: NTSA

Bohemia Interactive Australia (2008). Virtual Battlespace 2 Minimum system requirements webpage. Retrieved 18 December, 2009 from http://virtualbattlespace.vbs2.com/index.php?option=com_content&task=view&id=76&Itemid=74.

Caspian Learning (2008). Serious games in defence education. Report produced for UK Ministry of Defence. Accessed 20 April 2009 from www.caspianlearning.co.uk/MoD_Defence_Academy_Serious_games_Report_04.11.08.pdf

Colton, G. (2008). Enhancing operational capability: Making infantry more deployable. *Australian Army Journal, V (1),* 51-64.

Curtis, N., & Hobbs, W.S.R. (1997). *Characterisation of infantry section and platoon activities. (DSTO-TR-0482).* Salisbury, SA: DSTO.

De Freitas, S. (2006). *Learning in immersive worlds: A review of game-based learning.* Bristol, England: JISC. Accessed 23 January 2013 from http://www.jisc.ac.uk/media/documents/programmes/elearninginnovation/gamingreport_v3.pdf.

Fournier, H., Lapointe, J-F., Munteanu, C., Edmond, B., & Kondrovata, I. (2011) A multidisciplinary approach to enhancing infantry training through immersive technologies. *Proceedings of the 2011 Interservice/Industry Training, Simulation & Education Conference*. Orlando, Fl: NTSA

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.

Guille, M., & French, H. T. (2004). *Toward a methodology for evaluating the impact of technologies on infantry situation awareness (DSTO-TR-1628).* Edinburgh, SA: DSTO.

Hays, R.T. (2005). *The effectiveness of instructional games: A literature review and discussion (Technical Report 2005-004).* Orlando, FL: Naval Air Warfare Center Training Systems Division.

Hobbs, W., & Mouzakis, K. (2001). Investigation of pen-based hand held computing for combat soldiers. *Journal of Battlefield Technology, 4,* 28-33.

Hopkins, W. G. (2002). A scale of magnitude for effect size statistics. Retrieved 23 January, 2013, from http://www.sportsci.org/resource/stats/effectmag.html.

Hussain, T., & Feurzeig, W. (2008). Methods and tools for the development of effective training games, *American Educational Research Association Annual Meeting*. New York, NY.

Jarmasz, J., Muller-Gass, A., Zotov, V., Lamb, M., Scourtoudis, E., Wojtarowicz, D., Thomson, M., & Bruyn-Martin, L. (2010). Blended solutions for counter-IED training. Paper presented at the 2010 Spring Simulation Multiconference.

Jarvis, J. (2008). AI opening door [Msg 3]. Message posted to https://forums.bisimulations.com/viewtopic.php?f=5&t=1700.

Kirk, R. E. (2003). Experimental design. In I. B Weiner (Ed), *Handbook of Psychology* (pp1-32). New York: Wiley.

Knerr, B.W. (2006). Current issues in the use of virtual simulations for dismounted soldier training. In *Virtual Media for Military Applications* (pp. 21-1 to 21-12). Meeting Proceedings RTO-MP-HFM-136, Paper 21. Neuilly-sur-Seine, France: RTO.

Kneuper, G. M. (2006). *Substituting live training with virtual training by means of a Commercial Off The Shelf, First Person Shooter computer game and the effect on performance.* Unpublished Master's dissertation, the University of Central Florida, Orlando, FL.

Kulik, J.A. (1994). Meta-analytic studies of findings on computer based instruction. In E.L. Backer & H.F. O'Neill (Eds.) *Technology assessment in education and training* (pp. 9-33).

Law, V. (2008, September 8). In the driver's seat. *Army News,* p23.

Lewis, C. (2005). Validating a first person perspective game as a tool for soldier systems research. Unpublished Master's Thesis. Department of Psychology, University of Adelaide, South Australia.

Morrison, P., & Barlow, M. (2004). Child's play? Coercing a COTS game into a military experimentation tool. *Proceedings of the 2004 SimTecT Conference.* Sydney, Simulation Industry Association of Australia.

Morrison, P., Barlow, M., Bethel, G., & Clothier, S. (2005). Proficient soldier to skilled gamer: Training for COTS success. *Proceedings of the 2005 SimTecT Conference.* Sydney: Simulation Industry Association of Australia.

NHMRC (2007). *National Statement on Ethical Conduct in Human Research*. Canberra: Commonwealth of Australia.

Nolan, J. M., & Jones, J. M. (2005). *Games for training: Leveraging Commercial Off The Shelf multiplayer gaming software for infantry collective training.* Unpublished Master's thesis, Naval Postgraduate School, Monterey, California.

Padilla, J. J. (2012) Annex 2: Military Simulation Systems. In: *Engineering Principles of Combat Modeling and Distributed Simulation*. John Wiley & Sons, Inc. 851-868

Parchman, S. W., Ellis, J. A., Christinaz, D., & Vogel, M. (2000). An evaluation of three computer-based instructional strategies in basic electricity and electronics training. *Military Psychology, 12*(1), 73-87.

Pennell, R. M. (2003). *Dismounted infantry Virtual Environment training effectiveness trial (DIVE TET) 03 (Report TR030689).* Farnborough, Hampshire, UK: QinetiQ.

Prensky, M. (2001). *Digital game-based learning*. New York: McGraw-Hill.

Proctor, M. D., & Woodman, M. D. (2007). Training "Shoot House" tactics using a game. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology, 4,* 55-63.

Ratwani, K., Orvis, K. and Knerr, B. (2010) An evaluation of game-based training effectiveness: Context matters. *Proceedings of the 2010 Interservice/Industry Training, Simulation and Education Conference*. Orlando, Fl: NTSA.

Rencrantz, C. (2003). *Game based training and education: Testing the usefulness of game consoles in the Swedish Armed Forces (Report FOI-R – 1086 – SE).* Linköping, Sweden: Swedish Defence Research Agency.

Ricci, K. E., Salas, E., & Cannon-Bowers, J. (1996). Do computer-based games facilitate knowledge acquisition and retention? *Military Psychology, 8,* 295-307.

Robson, S. (2008). *Not playing around: Army to invest $50m in combat training games.* Accessed 23 January 2013 from http://www.stripes.com/article.asp?section=104&article=59009.

Roman, P. A., & Brown, D. (2008). Games – Just how serious are they? *Proceedings of the 2008 Interservice/Industry Training, Simulation and Education Conference.* Orlando, Fl: NTSA.

Smith, R. (2010). The Long History of Gaming in Military Training. *Simulation & Gaming, 41(1)*, 6-19.

Smith-Jentsch, K.A., Zeisig, R.L., Acton, B., & McPherson, J.A. (1998). Team dimensional training: A strategy for guided self-correction. In J. Cannon-Bowers & E. Salas (Eds.), *Making Decisions Under Stress: Implications for individual and team training* (pp. 271-298). Washington, DC: APA Press.

StatSoft, Inc. (2010). Electronic Statistics Textbook. Tulsa, OK: StatSoft. WEB: http://www.statsoft.com/textbook/.

Summers, J. (2012). Simulation-based Military Training: An Engineering Approach to Better Addressing Competing Environmental, Fiscal, and Security Concerns. *Washington Academy of Sciences, 98(1),* 9-30.

Tack, D.W., & Angel, H. (2005). *Cognitive Task analyses of information requirements in dismounted infantry operations. (CR-2005-057)*. Defence Research and Development Canada – Toronto, Canada.

Temby, P. (2005). *Using computer games for military training*. Paper presented at the 2005 Defence Human Factors Special Interest Group (DHFSIG) Meeting, DSTO Edinburgh, South Australia.

The Technical Cooperation Program (2006). *Guide for understanding and implementing defense experimentation (GUIDEx).* Ottawa, Canada: Author.

Topolski, R., et al. (2010) *Impact of game-based training on classroom learning outcomes (Study Report 2010-01)*. Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

van der Hulst, A., Muller, T., Besselink, S., Coetsier, D., & Roos, M. C. (2008). Bloody serious gaming – experiences with job oriented training. *Proceedings of the 2008 Interservice/Industry Training, Simulation and Education Conference*. Orlando, Fl: NTSA.

Whitney, S. J., Fidock, J. and Ferguson, N. (2012) Assessing the effectiveness of simulation-based counter-IED training. *Journal of Battlefield Technology, 15 (1)*, 57-64.

Whitney, S. J., Temby, P., & Stephens, A. (2008). Evaluating Game-Based Training: Criteria, Skills, and Future Directions. *Proceedings of the 2008 SimTecT Conference.* Sydney: Simulation Industry Association of Australia.

Whitney, S. J., Temby, P., & Stephens, A. (2010*).* Serious games need serious performance measures: A case study. *Proceedings of the 2010 SimTecT Conference*. Sydney: Simulation Industry Association of Australia.

Whitney, S.J., Temby, P., & Stephens, A. (in press). A review of the effectiveness of game-based training for dismounted soldiers *Journal of Defense Modeling and Simulation*.

Wiederhold, M. D. (2005). Physiological monitoring *during simulation training and testing (Report A851634)*. San Diego, CA: Virtual Reality Medical Center.

Woodman, M. D. (2006). *Cognitive training transfer using a personal computer-based game: A close quarters battle case study.* Unpublished Doctoral dissertation, the University of Central Florida, Orlando, FL.

Appendix A: Questionnaires

A.1. Background Questionnaire

1. Name: _____

2. Rank: _____

3. What is your current age? _____years

4. Years of Military Service _____years _____months

5. Have you ever experienced motion sickness?
   - ☐ Yes
   - ☐ No
   - ☐ Unsure

6. Have you ever experienced motion sickness while playing a computer game?
   - ☐ Yes
   - ☐ No
   - ☐ Unsure

7. Are you colour blind?
   - ☐ Yes
   - ☐ No
   - ☐ Unsure

8. Do you have normal or 20/20 vision?
   - ☐ Yes
   - ☐ No
   - ☐ Unsure

9. Do you own a computer?
   - ☐ Yes
   - ☐ No

10. How often do you play computer games?
    - ☐ Never
    - ☐ Daily
    - ☐ Weekly
    - ☐ Monthly

11. What types of computer games do you currently play (tick as many as appropriate)?
    - ☐ First person shooter games
    - ☐ Strategy games
    - ☐ Adventure games
    - ☐ Fantasy / D&D games
    - ☐ None

12. Are you right or left handed?
    - ☐ Right-handed
    - ☐ Left-handed
    - ☐ Both (ambidextrous)

13. Have you played VBS2 before?
    - ☐ Yes
    - ☐ No
    - ☐ Unsure

14. How many times have you played VBS2?
    - ☐ Never
    - ☐ 1-3 times
    - ☐ 4-6 times
    - ☐ 7-10 times
    - ☐ More than 10 times

15. Have you used other military computer simulations?
    - ☐ Yes If yes, which ones? _____
    - ☐ No
    - ☐ Unsure

16. Have you had training in section attack drills?
    - ☐ Yes
    - ☐ No
    - ☐ Unsure

17. Did you miss any training on infantry minor tactics?
    - ☐ Yes
    - ☐ No
    - ☐ Unsure

18. How confident are you in conducting section attack procedures?
    - ☐ Not at all confident
    - ☐ A little bit confident
    - ☐ Somewhat confident
    - ☐ Moderately confident
    - ☐ Very confident

A.1.1 Summary of Background Questionnaire Responses

Table 12: Summary of background questionnaire data (not included in body of report)

Question	No. of Responses by section	
Q5. Have you ever experienced motion sickness?	Game-based	Field-based
Yes	0	2
No	7	7
Unsure[a]	2	0
Q6. Have you ever experienced motion sickness while playing a computer game?		
Yes	0	0
No	8	9
Unsure	1	0
Q7. Are you colour blind?		
Yes	1	0
No	8	9
Q8. Do you have normal or 20/20 vision?		
Yes	7	8
No	1	0
Unsure	1	1
Q9. Do you own a computer?		
Yes	9	7
No	0	2
Q12. Are you right or left handed?		
Right	9	8
Left	0	1
Q13. Have you played VBS2 before?[b]		
Yes	0	0
No	9	8
Unsure	0	1
Q15. Have you used other military simulations?[c]		
Yes	7	6
No	2	3
Q16. Have you had training in section attack drills?		
Yes	9	9
No	0	0
Q17. Did you miss any training on infantry minor tactics?		
Yes	0	0
No	9	8
Unsure	0	1

Notes:

[a] The "unsure" category is included only for questions where one or more participants answered "unsure".

[b] As no participant reported experience with VBS2, all answered 'No' to Q14, which asks about frequency of VBS2 playing. Answers to Q14 are therefore not reported here.

[c] SOI staff observed that all participants had used the Weapons Training Simulation System; hence all participants should have answered "yes" to this question. It is possible that participants who answered "no" interpreted the question as referring only to desktop-based simulations.

A.2. Exit Questionnaire

Name: ……. **Date**: ……..

During the VBS2 training how easy was it to:

1. Navigate through the virtual environment?
   - ☐ Very Easy
   - ☐ Somewhat Easy
   - ☐ Somewhat Difficult
   - ☐ Very Difficult

2. Communicate with other section members?
   - ☐ Very Easy
   - ☐ Somewhat Easy
   - ☐ Somewhat Difficult
   - ☐ Very Difficult

3. Conduct fire and movement?
   - ☐ Very Easy
   - ☐ Somewhat Easy
   - ☐ Somewhat Difficult
   - ☐ Very Difficult

4. Maintain situation awareness of your team members' positions?
   - ☐ Very Easy
   - ☐ Somewhat Easy
   - ☐ Somewhat Difficult
   - ☐ Very Difficult

5. Identify the OPFOR?
   - ☐ Very Easy
   - ☐ Somewhat Easy
   - ☐ Somewhat Difficult
   - ☐ Very Difficult
   - ☐ Not Applicable

6. Achieve the mission objectives?
   - ☐ Very Easy
   - ☐ Somewhat Easy
   - ☐ Somewhat Difficult
   - ☐ Very Difficult

7. List 3 things you liked **best** about using VBS2 (and your reasons why)?

1.……..…
……....
2.…..……...…
……..
3.……...…
……..

8. List 3 things you liked **least** about using VBS2 (and your reasons why)?

1.……..…
……....
2.…..……...…
……..
3.……...…
……..

9. Was the **training time** you received on VBS2 adequate? Explain your answer.

……
……
……

10. What **modifications** would you make to **improve** VBS2 as a training tool?

……
……
……

11. How **effective** was the VBS2 training in preparing you for the section attack missions you conducted in the field? Tick one option and explain your answer.
    - ☐ Not at all effective
    - ☐ Somewhat effective
    - ☐ Moderately effective
    - ☐ Very effective

Appendix B: Additional details relating to game-based training

This appendix describes in more detail the technical details of the game-based training. The hardware and software components of the computers used in game-based training are summarised in Table 11.

Table 13: Hardware and software components used during game-based training

Component	Specifications
CPU	Pentium 4, 3.4 GHz
Operating System	Windows XP 32-bit, Service Pack 2
Memory	1 Gigabyte RAM
Video/Graphics Card	NVIDIA Geforce 7900 GTX
Audio	Legacy audio; stereo sound (non-3D)
Software Version	VBS2 VTK Release 1.4 - BIA 1.20.5634
Monitors	Single screen 19" LCD (BenQ FP92G+)

Figure 16: BLUEFOR avatar (left) and OPFOR avatar (right)

Within VBS2, BLUEFOR avatars had the appearance of Australian Army soldiers with respect to dress, as shown in Figure 16 (left). The avatars carried the standard weapons, ammunition, and equipment of the Australian Army, namely the F88 Steyr, the F89 Minimi, 9mm Browning, LAW, ammunition for these, smoke grenades, hand grenades, Claymore mines, and night vision goggles. Allocation of weapons, ammunition, and equipment was consistent with Army's current doctrine.

The OPFOR avatars were selected from the VBS2 OPFOR library. An example of an OPFOR avatar is shown in Figure 16 (right). The exact type of OPFOR avatar (e.g. Machine Gunner, Rifleman) used in each mission is included in the mission descriptions in the following sections.

B.1. VBS2 Scenarios

A practice session and four missions were developed for use in game-based training. The sequence in which the missions were run is in Table 12, noting that Mission 2 was run three times.

Table 14: Running order of game-based training missions

Running order	Mission number
1st	1
2nd	2
3rd	2
4th	3
5th	2
6th	4

Specific details of each mission follow.

B.1.1 Practice Session

OPFOR and BLUEFOR were inserted at opposite ends of a small island. There was no mission objective. Rather, this was an opportunity for participants to become familiar with the controls and weaponry. OPFOR comprised two Riflemen, and one Officer. A map of the island is shown in Figure 17.

Figure 17: Map of the island used in the practice session

B.1.2 Mission 1

Three OPFOR – one Officer and two Riflemen – manned a checkpoint on a road, as shown in Figure 18. BLUEFOR patrolled along this road, and cleared all OPFOR. A map of this mission is shown in Figure 19.

Figure 18 and all other images showing 3D views were taken from AAR files. Symbols such as the 'hostile' symbol above the OPFOR, the 'friendly' symbol over BLUEFOR, and the lines connecting section members are visible in AAR playback, but are not visible during play.

Figure 18: Screenshot of the checkpoint in Mission 1

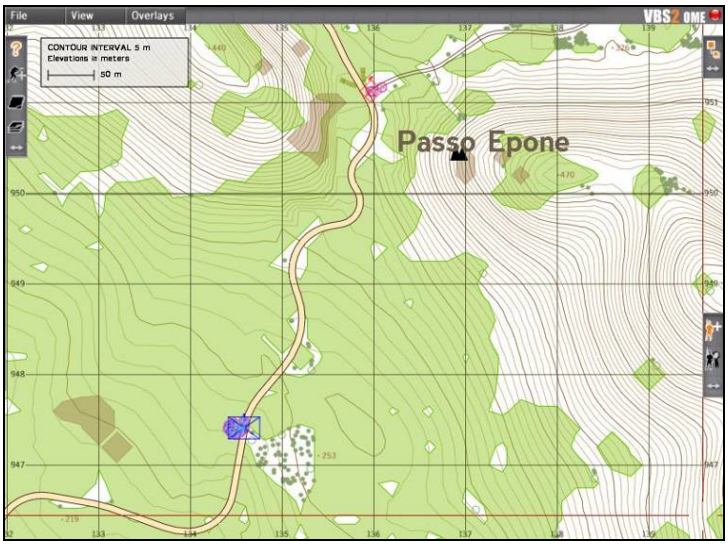

Figure 19: Map of Mission 1

B.1.3 Mission 2

This was run as the 2nd, 3rd, and 5th training mission. Two OPFOR Riflemen occupied a small group of buildings, as shown in Figure 20. BLUEFOR patrolled from the east, as shown in Figure 21.

Figure 20: Screenshot of the OPFOR location in Mission 2

Figure 21: Map of Mission 2

B.1.4 Mission 3

This was run as the 4th training mission. Three OPFOR – one Grenadier and two Riflemen – patrolled in a small valley, as shown in Figure 22. BLUEFOR, patrolling from the south, encountered the OPFOR. Figure 23 shows the map for this mission.

Figure 22: Screenshot of the terrain in Mission 3

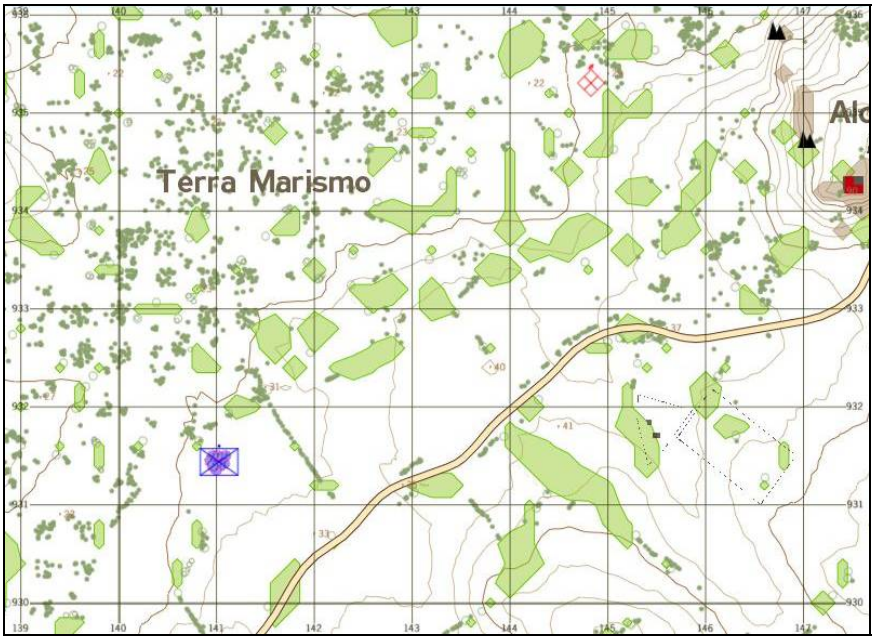

Figure 23: Map of Mission 3

B.1.5 Mission 4

This was run as the 6th training mission. This mission took place on a small island, as shown in Figure 24. Three OPFOR – two Riflemen and a Grenadier – were positioned among the buildings. The BLUEFOR starting position was on the bridge connecting the island to the mainland, as shown in Figure 25.

Figure 24: Screenshot of the terrain in Mission 4

Figure 25: Map of Mission 4

Appendix C: Knowledge-Based Assessment

As discussed in Section 3.4.3, the pre-training and post-training assessment data were reanalysed using an alternative rating scale where scores below 1 were given a rating of zero, and scores of 1 or more were given a rating of 1. A comparison of the pre-training and post-training assessment using this rating scale is shown in Figure 26. A Chi-square test based on the frequency of ratings of 0 and 1 confirmed that (1) there was no significant difference between the baseline performance of the two sections, $\chi^2 = 2.59$, $p = 0.11$ (df=1), however there was also no significant difference between the two sections on the post-training assessment, $\chi^2 = 1.91$, $p = 0.17$ (df=1). Chi-square tests also showed that both the game-based, $\chi^2 = 46.3$, $p < 0.001$ (df=1) and field-based, $\chi^2 = 31.1$, $p < 0.001$ (df=1) sections' performance increased significantly following training.

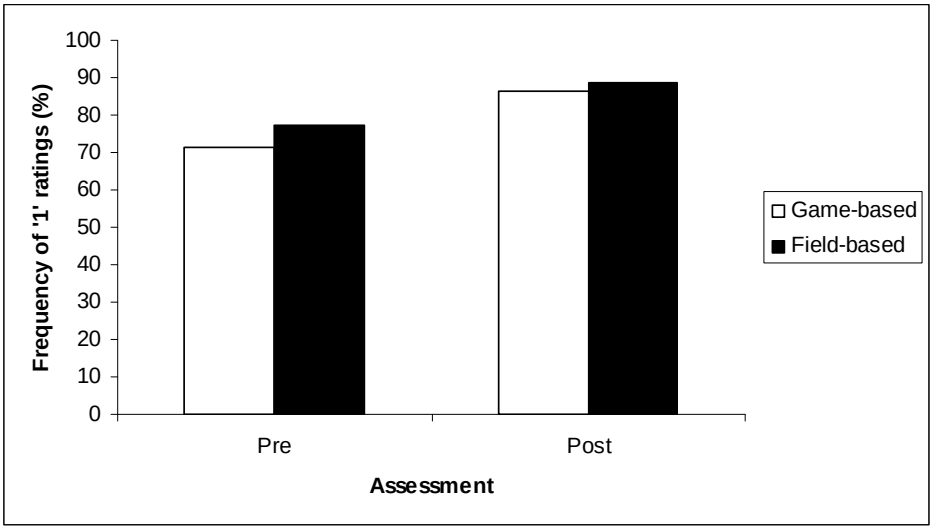

Figure 26: Comparison of average scores for pre- and post-training assessments using a knowledge-based rating scale.

Performance changes for each behaviour are shown in Table 15. For the game-based section, changes are significant for all behaviours. Significant performance increases occurred for the field-based training section for coordination and supporting behaviour.

Table 15: Frequency of ratings of 1 (as percentage of the total), χ^2 and p values (df =1) for pre- and post-training scores for both section.

	Game-based Section					Field-based Section				
	Pre	**Post**	**% Increase**	χ^2	***p*-value**	**Pre**	**Post**	**% Increase**	χ^2	***p*-value**
Communication	70.8	81.9	11.1	4.3	.04	83.3	88.9	5.6	1.6	.21
Leadership	83.3	100.0	16.7	14.4	< .001	87.5	98.6	11.1	8.1	.004
Coordination	56.3	80.2	24.0	22.4	< .001	59.4	83.3	24.0	22.8	< .001
Supporting Behaviour	50.0	66.7	16.7	8.0	.005	54.2	73.6	19.4	11.0	< .001
General Points	91.7	99.1	7.4	7.8	.005	97.2	96.3	-0.9	0.3	.56
Total	71.4	86.4	15.0	46.3	< .001	77.1	88.6	11.5	31.1	< .001

The results obtained after removing the 10 items deemed too difficult to conduct in VBS2 are presented in Table 16. A Chi-square test based on frequencies of 0 and 1 pre- and post-testing showed that both the game-based and field-based training sections' performance still increased significantly following training. When broken down by behaviour, the only significant performance change for the game-based training section occurs with coordination. This change was also noteworthy (i.e., greater than 10%). For the field-based section, performance increases were statistically significant for leadership, coordination and supporting behaviour. The magnitude of these changes is also noteworthy.

Table 16: Frequency of ratings of 1 (as percentage of the total), χ^2 and p values (df =1) for pre- and post-training scores after removing 10 items from assessment criteria.

	Game-based Section					**Field-based Section**				
	Pre	**Post**	**% Increase**	χ^2	***p*-value**	**Pre**	**Post**	**% Increase**	χ^2	***p*-value**
Communication	93.8	95.8	2.0	0.4	.55	100.0	100.0	0.0	-	1
Leadership	95.0	100.0	5.0	3.2	.08	90.0	100.0	10.0	6.7	.01
Coordination	64.3	79.8	15.5	8.8	.003	67.9	85.7	17.8	12.3	.005
Supporting Behaviour	56.3	68.8	12.5	3.0	.08	62.5	79.2	16.7	5.7	.02
General Points	95.0	98.3	3.3	1.4	.24	100.0	95.0	-5.0	3.2	.08
Total	80.0	88.3	8.3	13.0	< .001	83.0	91.7	8.7	16.0	< .001

Appendix D: Analysis of TTCP GUIDEx Threats to a Good Defence Experiment

Table 17: List of GUIDEx threats to a good experiment and strategies used to mitigate them in the current study

Threats	Description of Threat	Mitigation Strategies Used in Study
1. Ability to use capability		
Capability not workable	The extent to which the technology being tested functioned properly. Do the hardware and software work?	Yes. Extensive pre-trial testing was undertaken to ensure that the hardware and software (VBS2) functioned properly. The technology used to run the game-based training functioned without any major problems for the duration of the study. The hardware used in the study all exceeded the minimum requirements for running the VBS2 software, although the resolution had to be set to the lowest level to prevent computer crashes from occurring. This setting did not appear to have any significant effect on the study outcomes.
Player non-use	The extent to which participants were familiar with the equipment used. Do players have the training and TTP to use the capability?	Yes. All participants were familiar with the weapons used in the study from their basic training. Participants in the game-based section were given adequate time to become familiar with the game controls. Both sections were trained in section attack procedures by experienced military personnel using current Australian Army doctrine.
Lack of potential effect in output	The extent to which independent variables are capable of producing measurable effects in the dependent variable. Is the output sensitive to capability use?	Yes. The training time was considered by the assessing staff to be adequate to produce performance improvement, and the training structure was similar to that currently used by the Australian Army.
Capability not exercised	This relates to whether the capability was able to be demonstrated as intended under representative trial conditions.	Yes. The capabilities of interest to the study, namely VBS2 and section attack, were both able to be exercised as intended during the study without any major problems.
2. Ability to detect change:		
Capability variability	Are systems (hardware and software) in use in like trials the same?	Yes. The computer hardware and software performed at a consistently stable level for the duration of the game-based training part of the study.
Player variability	Do individual operators/units in like trials have similar characteristics?	Yes. All participants had similar levels of military experience and were given refresher lessons in section attack. The baseline assessment confirmed that the skill levels of the two sections were comparable prior to undertaking training. Demographic data was similar for the two groups.

Data collection variability	Is there large error variability in the data collection process?	No. Data collection was undertaken by the same assessors for the pre- and post-training tests. All assessors were experienced in assessing section attack using the criteria outlined in Table 2. They discussed and calibrated performance ratings before the commencement of the study, which reduced the likelihood of variability and increased the reliability of their ratings.
Trial conditions variability	Are there uncontrolled or unmonitored changes in trial conditions for like trials?	No. The study conditions for the pre- and post-training assessments were virtually identical for the two groups with respect to factors such as starting location, OPFOR position, and mission objectives, and were conducted at similar times of day in similar weather conditions.
Low statistical power	Is the analysis efficient and the sample sufficient?	Yes. It is unlikely that low statistical power was a threat to this experiment, given the large number of data points used in the statistical analyses. Furthermore, moderate effect sizes were obtained in the study.
Violation of statistical assumptions	Are the correct analysis techniques used and error rates reduced?	Yes. Non-parametric tests were used for all data analyses, except for cases where the appropriate parametric test was known to be robust to deviations from normal distributions. Error rates were reduced by using multiple comparison tests (e.g. mixed ANOVA tests).
3. Ability to isolate reasons for change:		
Player differences between experimental conditions	Are there differences between groups unrelated to the treatment?	No. A stratified sampling method was employed to match the experimental and control groups on key demographic variables. The two groups were comparable in length of service, frequency of game playing, experience with VBS2, and confidence in conducting section attack. The pre-training assessment confirmed that the two sections were comparable in their section attack performance level prior to receiving training.
Data collection differences	Are there potential data collection differences between treatment groups?	No. The same SMEs were used to assess both treatment groups in the pre- and post-test trials, which reduced the likelihood of data collection differences between groups. In addition, two SMEs were blind to the type of training each group received, which reduces the likelihood of data collection bias. Blind and non-blind assessor scores were not statistically different.
Trial conditions differences	Are the trial conditions similar for each treatment group?	Yes. Both groups received comparable training on section attack during the study. This included similar amounts of training time, training sessions that were similar in structure, and comparable levels of mission difficulty. These conditions increase the

		likelihood that the differences in training outcomes are due to differences in training methods, rather than differences in training exposure.
4. Ability to relate results to operations:		
Non-representative capability	Is the experimental surrogate functionally representative?	Yes. This was not a threat in this experiment, as VBS2 is a mature capability. Furthermore, the section attack procedures used in the study have been used for many years by the Australian Army, and are unlikely to change significantly in the future.
Non-representative players	Is the player unit similar to the intended operational unit?	Yes. The participants were representative of IET trainees, to whom the capability (VBS2) might be of use. Novice infantry soldiers were an appropriate sample to use because it was expected that they would improve more from the training than would experienced soldiers.
Non-representative measures	Do the performance measures reflect the desired operational outcome?	Yes. The performance measures were the same measures used by the Australian Army's School of Infantry for assessing section attack, which have been developed over many years. The 6-point rating scale (with behavioural descriptors for each rating point) also provided sufficient granularity to detect genuine differences between groups with a high degree of certainty.
Non-representative scenarios	Are the blue, green, and red conditions realistic?	Yes. The scenarios used for the game-based and field-based activities were developed in conjunction with Army SMEs with over 20 years of experience. This input ensured that the scenarios were realistic, including a reactive OPFOR, and appropriate for the experience level of the participants.

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION
DOCUMENT CONTROL DATA

1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)

2. TITLE

Evaluating the Effectiveness of Game-Based Training: A Controlled Study with Dismounted Infantry Teams

3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION)

Document	(U)
Title	(U)
Abstract	(U)

4. AUTHOR(S)

Susannah J. Whitney, Philip Temby, and Ashley Stephens

5. CORPORATE AUTHOR

DSTO Defence Science and Technology Organisation
PO Box 1500
Edinburgh South Australia 5111 Australia

6a. DSTO NUMBER	6b. AR NUMBER	6c. TYPE OF REPORT	7. DOCUMENT DATE
DSTO-TR-2799	AR-015-508	Technical Report	January 2013

8. FILE NUMBER	9. TASK NUMBER	10. TASK SPONSOR	11. NO. OF PAGES	12. NO. OF REFERENCES
2010/1119536/1	ARM 07/163	Forces Command	61	57

DSTO Publications Repository

http://dspace.dsto.defence.gov.au/dspace

14. RELEASE AUTHORITY

Chief, Land Operations Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for public release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111

16. DELIBERATE ANNOUNCEMENT

No Limitations

17. CITATION IN OTHER DOCUMENTS Yes

18. DSTO RESEARCH LIBRARY THESAURUS

Infantry, military training, simulation, team performance, training evaluation

19. ABSTRACT

Computer games are increasingly being used by armed forces to supplement traditional methods of military training, despite a lack of empirical evidence on their training effectiveness. This report describes a study conducted by DSTO scientists examining the effectivensss of a desktop computer game to train small teams of dismounted soldiers in infantry tactics, techniques, and procedures. One infantry section received traditional field-based instruction in section attack procedures, the other took part in game-based training using Virtual Battlespace 2. The performance of both sections was measured before, during, and after training. While the performance of the field-based training section improved significantly from pre-training to post-training, the game-based section showed no significant changes in performance. Overall the findings suggest that the current method of field-based training is effective, and that game-based training is not effective for training novice teams of infantry personnel in section attack procedures. The implications for Defence for using desktop computer games for individual and team training are discussed and recommendations for future research into game-based training are outlined.