Factors in Team Performance in a Virtual Squad Environment

Michael Barlow[¥]; Mathew Luck[¥]; Edward Lewis[¥]; Mike Ford[¥]; Robert Cox[±]

(¥) Virtual Environments & Simulation Laboratory (VESL)
School of ITEE
UNSW @ ADFA
{spike,m.luck,e.lewis,m.ford}@adfa.edu.au

(±) School of Computer Science
University of Canberra
robert.cox@canberra.edu.au

Abstract. COTS computer games have been receiving increasing interest as to their potential application in the military sphere – ranging from training through experimentation, and decision-support, and even recruitment. This paper describes an experiment using the Operation Flashpoint 1st person Shooter game with small teams in a competition. Major factors investigated were which factors contributed to the success of a team in a human-on-human environment and the decision making and organisation process employed by the participants. The paper describes the methodology of the experiment, the results of the examination of the decision process and the team performance results. In the discussion factors are considered that may have implications for the use of 1PS games for military purposes and some thoughts given as to how this experimentation can be improved and continued.

1. INTRODUCTION

COTS (Commercial Off The Shelf) computer games have been receiving increasing interest as to their potential application in the military sphere – ranging from training [2],[8] through experimentation [4], and decision-support [3], and even recruitment [11]. COTS computer games, particularly 1st Person Shooters (hereafter referred to as 1PS), are believed to possess considerable, as yet unrealised, potential because of their ability to represent combat realistically for teams of players, in an way that can be readily adapted for analytical purposes. If this potential is realised then this cheap tool could provide a valuable supplement to the tactical engagement simulation advocated by Manton et al [9].

On the other hand there are risks in the military application of COTS games which stem from the fact that the original product was built for entertainment rather than simulation veracity; while other risks are simply a consequence of a relatively recent and dynamic technology being used in a novel fashion.

This paper describes an experiment that seeks to illuminate several of these risks in the use of a 1PS for military purposes. A major focus of the experiment was to explore the use of a 1PS in a competitive team environment and what factors correlate with the success or failure of a team. This exploration is of particular importance when a 1PS is used for experimentation and decision support – for example doctrine development or evaluation, or the evaluation of a new capability (e.g., equipment). In order to establish the validity of the simulation for such applications, it is vital to know what factors external to the simulated environment – for instance prior experience with the 1PS, computer gaming familiarity, degree of communication afforded the participants, etc. – will impact the outcome of experiments and to what extent.

These experiments can also be used to examine the decision-making and organisation process employed by the participants – the players on each team – to see if these games can be used to understand, or improve through practice, these processes. In particular, the question is asked whether participants are following a Naturalistic Decision Making [7] or heuristic [5] process during scenario participation. NDM is a theory about the way humans make decisions under the constraints of incomplete knowledge, time pressures, a dynamic environment, and high costs for errors; and appears to encapsulate many situations faced by a military commander. As the NDM process is at the heart of the military decision-making process taught to Army battle commanders [1],[6], it would be useful to see if this process is as natural and as useful as advocated by Klein. If this is true, a 1PS could be used to provide the critical decision-making experience that is necessary to NDM and enhance critical thinking skills for battle command [10]. Examining how a team self-organises – whether centralized command, or breaking into sub-units, or fixed roles – and how these correlate with team performance is a first step in pursuing the possibility of the use of a 1PS for doctrine development.

2. METHODOLOGY

A round-robin team competition was constructed using version 1.90 of the game Operation Flashpoint. Operation Flashpoint was selected due to its fidelity for modern military operations, and high degree of configurability [2],[3].

Matches in the competition were conducted using two separate rooms, with each team in their own room. Both rooms possessed modern PCs with powerful graphics cards. Each computer was equipped with a headset, and players of the same team were free to speak to one another for in-game communication.

A separate machine (to the players') was employed to serve all matches; ensuring that no player's machine was excessively burdened with controlling the total game state.

The experiment was organised as a competition[1] that was open to all staff and students at the Australian Defence Force Academy (ADFA). Participants formed their own teams of five plus reserves. In all eight (8) teams were formed for a total of forty six (46) participants. Seven of the teams were composed of military personnel – five teams were composed of officer cadets and midshipmen, and two teams were composed of officers undertaking undergraduate studies at ADFA. The final team was civilian, composed of staff of one of the schools at ADFA.

A competition draw was arranged as seven rounds of competition – such that each team played all of the others. A ladder was maintained of team positions in the competition – playing in a match earned a team one point, while each rubber they won (of the three that composed a match) earned the winner three points. Hence a team could earn as much as ten points for a match.

Each match between two teams comprised three rubbers of twenty minutes duration each; for a total match time of 1 hour. Between each rubber there was a five minute break. The same scenario was used for all three rubbers that comprised a match. However, for each rubber teams swapped sides in the scenario being played (all scenarios had two sides designated East and West) – analogous to teams swapping sides at half-time in a football match. The third rubber differed in that the penalty for death was altered (increased), as explained in the next subsection.

Four different team-based scenarios were used for the competition – with two being specially constructed for the competition. In brief the scenarios were Capture And Hold (CAH), Capture The Flag (CTF), Proximity, and Forcings Back (FB). CAH comprised a central region (square, courtyard, etc.) of which only one team could be in control at any time – that team would earn a point every second. At the end of twenty minutes the team with the most points won the rubber (this applies to all four scenario types). CTF is played with each team possessing a flag that is held close to their base. Teams earn points by successfully capturing the enemy team's flag and delivering it to their base (while stopping the removal of their own flag). Proximity is played with a 'beacon' situated in the centre of the map. Teams accumulate points on a second-by-second basis based on the proximity (distance) of all their members to the beacon. FB is also scored on a second-by-second basis. Teams seek to drive the opposed side back to their base while advancing steadily and securely. Advancing too far (behind the enemy frontline) is penalised. In all, five different map terrains were used for the seven different scenarios. Most were urban style environments with enterable one and two-story buildings and relatively short engagement ranges.

All rubbers were played with respawning – participants killed in the simulation were not removed from the rubber, but after a suitable delay were reincarnated (respawned) at their base. A pilot study in which no respawn was employed found that all match types, regardless of scenario, 'degenerated' into a 'deathmatch' – kill the enemy regardless of victory condition. For this reason, and in order to maintain the focus of the work on team operations by preserving a full team throughout a rubber, respawn was employed in all matches and rubbers. A respawn delay – time between death and reincarnation – of 5 seconds in the first two rubbers, and 30 seconds for the third rubber was employed.

Each soldier in the simulated VE (game) was equipped with an assault rifle (Steyr when playing East and G36 when playing West) with 1.5x scope and four magazines of ammunition. They also carried three hand grenades and two smoke grenades. Each time a player respawned they were provided with the same equipment as they began the rubber with.

Participants in the experiment were surveyed prior to, and after completion of the competition, as well as following each match they played. Prior to competition start participants were also measured as to their 'coordination' in the VE. Analysis of those measurements and survey responses forms much of the subsequent data.

The enrollment survey quantified participant's computer gaming background, amount of time spent playing computer (or console) games per week, familiarity with OFP (Operation Flashpoint) and their military background (if any).

[1] Known as the BIA VESL Cup – See http://www.cs.adfa.edu.au/research/vesl/OFPcomp

The pre-competition coordination test sought to measure participant basic skill in the game space. It consisted of four trials somewhat similar to a boot camp. An obstacle course that included movement between designated stations, for which completion time was recorded. A firing range, for which time to hit all targets was recorded. A grenade range, for which time to 'hit' all targets was recorded. And a 'sneaker' range in which the participant had to move and fire at targets that periodically popped-up, for which completion time and number of targets hit was recorded.

Following each match in which they played participants filled out a questionnaire that asked them what tactics they used and what was effective (and what was not), what fraction of the time they worked directly with teammates, how much verbal communication was used (and what was communicated), their perceptions of the pace of each rubber, what fraction of the time they were aware of teammate and enemy positions, and how often they used different criteria for making play decisions.

The post-competition survey asked participants as to their evaluation of OFP as a military tool (same question set as [2]), their perception of and view of the four different game types, the importance of a number of different factors (individual skills, communication, knowledge of friendly and enemy positions, etc.) in team performance, what criteria and how often they applied those criteria in choosing individual tactics across the length of the competition, their evaluation of their own and their team change in skill and tactics across the length of the competition, and how and to what extent changing the death penalty in a rubber altered their play.

3. ON PLAYER DECISION-MAKING IN A 1PS

The after match questionnaire, amongst other things, asked players how they made individual tactical decisions (e.g., to move to a piece of cover and cover an area from there, to aggressively attack around the opposed team's base). They were given three categories – 'considered evaluation of all possibilities', 'intuition/gut feel', and 'role (assigned/ordered/requested)'. For each they were to provide an adjectival rating from the set {never, rarely, occasionally, uncertain, often, almost always, always}. Figure 1 summarises the two hundred odd responses received.

Table 1: Statistical summary of method of decision making employed by participants.

	Considered	Intuitive	Assigned Role
Mode (Quart. rng.)	-0.3 (1.6)	0.7 (1.4)	1.2 (1.4)
Modal Adjectival Range	Often	Uncertain – almost always	Often – almost always

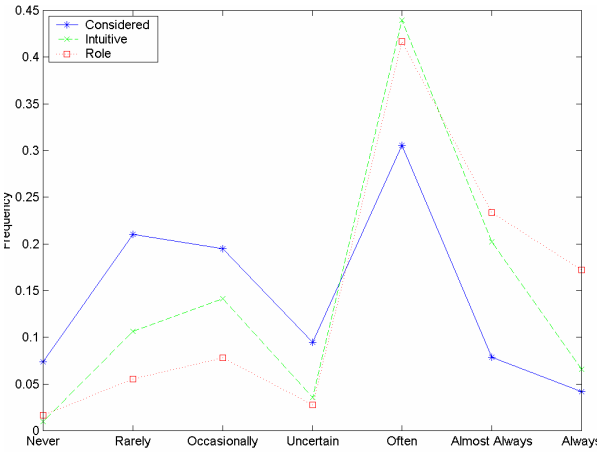

Figure 1: Mode of decision making used by participants when choosing to undertake actions during a match. Horizontal axis is participant ordinal response, vertical axis is the frequency of response. Participants provided responses for considered, intuitive, and dedicated (assigned) role.

Mapping the adjectival ordinal range to the numeric range –3 (never) to 3 (always); allows the calculation of statistics as seen in table 1.

Both the figure and table show that a range of decision making modes were employed, with assigned role the most heavily used (mode of 1.4), followed by intuition (0.7), and considered evaluation (-0.3).

4. FACTORS IN TEAM PERFORMANCE

The final competition point total for a team was interpreted as a direct measure of the performance of that team. The eight point values (ranging from 52 down to 20) for the eight different teams could then be correlated with other measurements or statistics available on each team (each team's players). This allowed analysis of which factors 'predicted' team outcome, and hence which might impact outcome of simulations using a 1PS for military decision making.

Upon enrolment, participants were asked as to their gaming background and habits – the number of years they had played computer games (and various genres of computer games), the number of hours per week they spent playing games, how many 1PS they had played/finished and the number of hours they had spent playing the 1PS that would be used for the

experiment (Operation Flashpoint). Those values were then averaged on a team basis and those averages (per team) then correlated against the points earned by the team. Table 2 summarises those correlations. High (absolute) values indicate strong predictors of team performance – variety in 1PS played and hours played per week, being more important than experience with OFP or years of gaming.

Table 2: Correlation of gaming background of players in a team with team performance.

Years of Gaming	Hours per week gaming	Variety of 1PS Team Games	# 1PS Played	Hours spent playing OFP
0.29	0.40	0.55	0.53	0.16

The results of the initial objective test were averaged on a team basis and correlated with the team's final performance. Table 3 presents these results with high values (in an absolute sense) indicating a strong predictor. Clearly coordination, particularly the weapon skills (e.g., Firing at –0.70 and grenades at –0.85), was a significant predictor.

Following each match in which they played, participants were surveyed on a number of factors concerning the match in question. These were then aggregated across all matches on a per-team basis and the values correlated against each team's final point tally. Surveyed factors included the amount of time working directly with teammates, perception of the pace of the match, portion of time aware of the disposition of teammates and enemy, and the decision making process employed. Table 4 shows the correlation of these factors with team performance. Significant predictors of team performance were found to be the amount of verbal communication used by a team (0.83), the use of roles within the team (0.58), and awareness of the disposition of own-team units.

Figure 2 is a plot of team points against each team's mean score for the coordination test, each team's mean verbal communication level for their matches, and finally a normalized (into the range 0-1, and ascending) equi-weighted sum of coordination and verbal communication level. The sub-plots show the degree of variability in these factors as predictors of team performance. Table 5 further analyses the post-match survey results by aggregating on the basis of whether the response came from a match winner or loser. Significant differences are seen in teamwork (sub-units of 3 and up), awareness of own team's disposition, and use of roles to manage the team.

At the end of the competition, participants were surveyed as to what factors they believed were most important for team victory. Five factors were presented and respondents provided each with a score from 1 (not significant) to 10 (most significant). The mean and standard deviation for each is found in Table 6. Team communication was seen as the most important (9.3) with individual skills, and situational awareness also being rated as very important (8.0 – 8.3).

5. 1PS SIMULATION FACTORS

Several other measured factors are mentioned briefly (due to space constraints) here.

The 3rd rubber of each match was played with a different death penalty – a wait of 30 seconds for a participant to re-enter the scenario, as opposed to 5 seconds for the first two rubbers. Participants were surveyed as to what extent the different (greater) penalty altered their play. The average response for movement tempo, use of cover, personal tactics, and team tactics indicated between minor and significant change while firing procedure was only a minor change. Such results indicate that changing the penalty for a game death (5 second wait versus 30 second wait) can significantly alter a player's approach to the simulation.

Participants were also surveyed as to their personal and team skill change across the course of the competition. The average response indicated that participants believed their individual coordination and personal tactics both underwent a minor improvement, while their 'perception/awareness' (ability to 'read' the scenario state/situation) underwent somewhere between a minor and serious improvement. Team-wise participants believed that their teamwork and team planning both made somewhat between minor and serious improvement.

Table 3: Correlation of 'coordination within 1PS' of players in a team with how the team performed.

Movement	Firing	Grenades	Sneaker	Normalised coordination
-0.30	-0.70	-0.85	-0.24	-0.78

Table 4: Correlation of post-match measures with a team's overall performance.

Working with 1 friendly	Working with 2+ friendly	Amount verbal comms	Percept of pace	Aware 1+ friendly	Aware 2+ friendly	Aware 1+ enemy	Aware 2+ enemy	Consider	Intuit	Role
-0.22	0.29	0.83	0.44	0.52	0.31	0.10	0.13	0.01	0.29	0.58

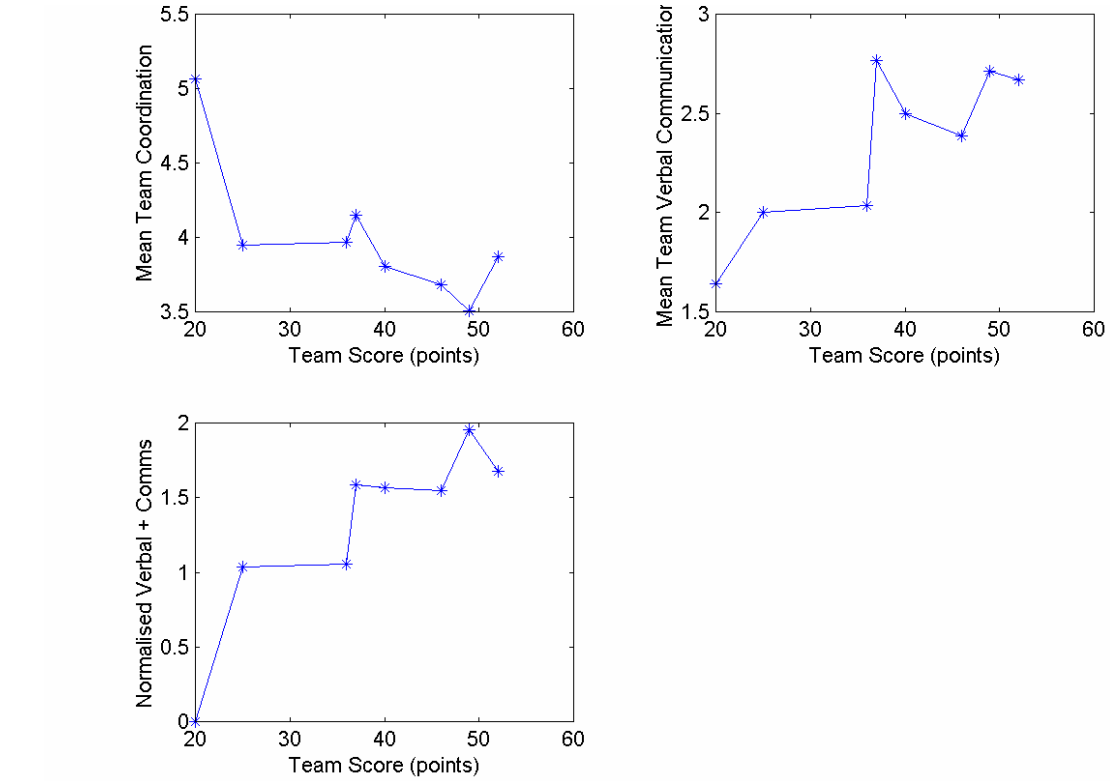

Figure 2: Plot of team performance (score) against average team performance on coordination tests [top-left], average level (amount) of verbal communication by team members during matches [top-right], and a linear-sum of the normalized (into 0-1 ascending range) of the two [bottom-left].

Table 5: Post-match survey responses from participants aggregated on basis of win/lose.

	Working with 1 friendly	Working with 2+ friendly	Amount verbal comms	Percept of pace	Aware of 1+ friendly	Aware of 2+ friendly	Aware 1+ enemies	Aware 2+ enemies	Consider	Intuit	Role
win	0.38	0.34	2.71	7.61	2.23	1.25	0.69	-0.46	-0.15	0.68	1.48
Lose	0.33	0.17	2.11	7.73	1.70	0.77	0.55	-0.78	-0.36	0.64	0.81

Table 6: Mean (std. dev. in brackets) of post-competition survey responses from participants about which factors they believed were most important for team victory.

Individual Skills	Aware team Disposition	Aware enemy Disposition	Team Communication	Familiarity with Scenarios
8.3 (1.4)	8.2 (1.1)	8.0 (1.2)	9.3 (0.9)	7.1 (1.9)

An analysis of the casualties in each rubber revealed a number of facts. There were an average of 120 deaths per 20-minute rubber – matching well with participant's perceptions that the pace of rubbers was very high (mean of 7.7 on scale of 1-10). There was a correlation (of 0.75) between difference in number of casualties suffered and difference in number of points earned – the winning side also tended to inflict more casualties on the opposition; though this was not always the case. An analysis of the source of casualties indicated that an average of somewhere between 10 and 20% of all casualties suffered (dependent on scenario type and teams involved) were not the result of enemy fire. In other words 'friendly fire' was a very real and frequent phenomenon in the games.

Finally contrasting the 4 different scenario types significant differences were found in casualty rates, perceptions of pace, amount of direct teamwork exercised, knowledge of friendly and enemy disposition, and decision-making processes. This tends to indicate that scenarios can be custom designed as to their rule-set (scoring system and rewards and penalties) to encourage and discourage different styles of play.

6. DISCUSSION

The results concerning decision-making in the 1PS indicate that participants employ a mixture of approaches – deliberate considered, intuition, and assigned role; with assigned role and intuition being more dominant. More successful teams employed role assignment and intuition more often than unsuccessful teams. The results of the players' knowledge of the disposition of friendly and enemy forces show that they rarely have complete knowledge of the disposition of their own forces, little alone that of the enemy. This, combined with the participants' perception of the pace of the rubbers (very high – dynamic), the high cost of mistakes (out of game temporarily and potential loss of match for team) describe the key conditions of an NDM environment. When this is combined with the data regarding the decision-making processes of the participants – the strong use of intuition as well as assigned roles to reduce the complexity - the proposition that a 1PS is a good vehicle for the investigation of NDM theory, seems well founded.

Analysing the factors correlated with team performance; certain relationships are clear. A background, or considerable experience (as measured by years of play or hours per week) was a contributing factor; though not as significant as the authors had expected. Even prior experience with the particular 1PS was only weakly correlated. However, actual coordination within the VE – ability to move and shoot rapidly – was a significant predictor of team performance. This is not surprising; particularly as the scenarios emphasised urban or close cover environments – with the parallel reduction in engagement ranges and visibility. Level of team communication was also found as a significant predictor of performance – those that communicated more tended to perform better. This is no doubt attributable to such communication providing greater situational awareness (a result borne out by the awareness values in table 5) and a common operating picture. Similarly, those teams that used role assignment more heavily tended to perform better – perhaps because the use of roles helped to simplify the inherent complexity of the VE and scenario.

The results on performance have implications for the use of a 1PS for military decision-support or training. For decision-support, such as experimentation [4] efforts should be made to normalise – such as by initial familiarisation, and training for competency - for factors such as inequalities in experience or skill with the technology of 1PS. Similarly, other factors that might influence outcome (doctrine, communication 'bandwidth') should be carefully controlled. From a training perspective a key enabler in the successful employment of 1PS would appear to be a decoupling of success with the *interface* (e.g., ability to position mouse cursor so as to make a successful shot) of the 1PS from success in the training task (e.g., acquisition of doctrine) itself.

The methodology of constructing an experiment at the core of which lies a competition (The BIA VESL Cup) in which military personnel compete in simulated military environments has considerable potential for yielding further insights – not only into

issues in the use of such technology but even as a vehicle for the investigation of questions of direct interest to the military (e.g., doctrine development, or capability evaluation) and areas of human decision-making. Plans for the 2004 competition include a comprehensive logging of all events (positions and facings, shots fired, fall of shot, hits and locations) within each rubber and an analysis of such data, new scenarios (tasks) that provide wider choices for individuals and teams, the modeling of round suppression and how that alters outcomes, measurement of the physiological arousal of participants and an analysis as to how that affects decision-making; as well as refinements in a number of other areas of the process. Details can be found at the VESL webpage – http://www.cs.adfa.edu.au/research/vesl.

ACKNOWLEDGEMENTS

The authors would like to acknowledge and thank the participants in the experiment.

REFERENCES

1. Australian Defence Force (1999), *Joint Publication 9 – Joint Planning*, Commonwealth of Australia, Canberra, Australia
2. Barlow M.; Morrison P. & Easton A. (2002) "1st Person Tactical Shooters: COTS Games with Military Training Potential", *Proceedings of SimTect2002*, Melbourne, May 2002.
3. Barlow M., Ford M., Lewis E. (2002) "The Use of Commercial Off the Shelf Games by Military Forces", Report for Australian Defence Simulations Office, June
4. Barlow M.; Morrison P.; Luck M. & Dickie A. (2004) "Constructing the Virtual Section", *in Proceedings*.
5. Gigerenzer, G. (2001) *The Adaptive Toolbox*, in Gigerenzer. and Sleten, (2001), 37-50
6. Greenhill, S., Venkatesh, S., Pearce, A., and Ly, T. (2002) Representations and processes in decision modelling, *DSTO-GD-0318*, Defence Science And Technology Organisation, Aeronautical and Maritime Research Laboratory, Fishermans Bend, Australia
7. Klein, G. (1989) Strategies for Decision making. *Military Review*, May, 56-64
8. Macedonea M. (2002), "Games Soldiers Play", *IEEE Spectrum OnLine*, 28 Feb 02
9. Manton, J., Wilson, Carlene, and Braithwaite, Helen (2000) "Human factors in field training for battle: realistically repres4enting chaos" Ch 11 in Evans and Ryan (2002), p177-197
10. Riedel, Sharon, Morath, R. and McGonigle, T. (2001) Training Critical Thinking Skills for Battle Command, *Report 1777*, US Army Research Institute, Alexandria, VA
11. Zyda M., Hiles J., Mayberry A., Wardynski C., Capps M., Osborn B., Shilling R., Robaszewski M., Davis M. (2003) "The MOVES Institute's Army Game Project: Entertainment R&D for Defence", *IEEE Computer Graphics and Applications*, Jan/Feb