Freeman, J., MacMillan, J., Haimson, C., Weil, S., and Diedrich, F. (2005). Systems, studies, and strategies in game-based learning. Proceedings of Training & Simulation International. 22-24 March, 2005. Maastricht, Netherlands.

SYSTEMS, STUDIES, AND STRATEGIES IN GAME-BASED TRAINING

Jared Freeman, Ph.D., Jean MacMillan, Ph.D.,
Craig Haimson, Ph.D., Shawn A. Weil, Ph.D., & Fred Diedrich, Ph.D.
Aptima, Inc.
Boston, MA & Washington, DC
United States

Online games deliver engaging opportunities for practice. Training, however, is more than practice; it is structured practice with feedback. To structure practice requires training objectives and control of the practice environment. To provide feedback requires measurement and assessment of human behavior. Further, a full training system should support scenario authoring, real-time monitoring and control of training, and post-practice training evaluation and administration. Current games offer little or no support for many of these core training functions. We describe the place of online games as practice environments within the larger training system, several specific gaming environments we are applying to training, and pilot studies that illustrate the strengths and weaknesses of online games as training vehicles.

INTRODUCTION

Online games deliver what much training lacks: experience so compelling it sells itself. Commercial sales figures make the case well. Sales of online games in the U.S. rose 167% in 2003 to $1 billion (Mercury News, 2004). Sales in China rose 64% to $157 million the same year (IT Facts, 2004).

Many of these games – notably first person shooters, strategy games, and flight simulators – simulate military operations. Thus, it is not surprising that the defense community has shown an increasing interest in leveraging these technologies for military training.

What, precisely, is the enticing experience that online games offer to players? It is the opportunity to immerse oneself within a visually captivating environment and to exercise one's skills at a rapid pace. In short, games offer *practice* in cue recognition and response[1].

However, training is more than practice. We view training as a system consisting of humans and technology, interacting through input and output relations. Humans serve several roles in this system: trainees, trainers, training controllers, observers, and evaluators or researchers. Training technologies include simulation environments (e.g., a game space), behavioral models (Artificial Intelligence (AI) or NonPlayer Characters (NPCs)), Learning Management Systems (LMSs), scenario authoring tools, measurement systems, feedback systems, and other tools.

Each of these entities has distinct relations to the others (which we represent using the notation $H_{entity} \leftarrow\rightarrow T_{component}$ to denote the flow of control or information ($\leftarrow \rightarrow$) between a human entity (H) such as a trainer, and a technology component (T) such as an authoring system). For example, a measurement system may assess trainee (in)actions ($T_{measurement} \rightarrow H_{trainee}$), trainees may assess themselves and their teammates ($H_{trainee} \rightarrow H_{trainee}$), trainees may evaluate the simulation itself ($H_{trainee} \rightarrow T_{simulation}$), and the simulation may assess the health of its own components during training ($T_{simulator} \rightarrow T_{simulator}$).

Figure 1 is a simplified representation of the many interactions between human users and training technologies. Some of these interactions are fulfilled by online games. Many are not.

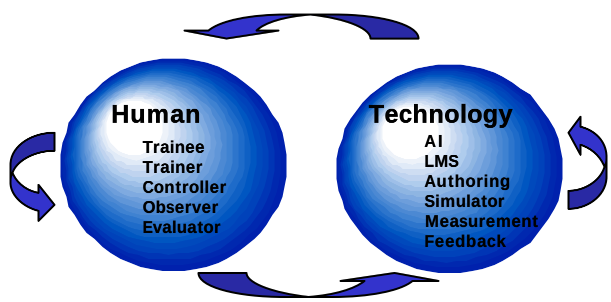

Figure 1: Training as a system of interactions within and between human users and technologies .

We are exploring the role of gaming simulations in training systems in several recent projects for the Defense Advanced Research Projects Agency (DARPA), the U.S. Army, and the U.S. Navy. In the

[1] Note that many games do not entirely satisfy the requirement of visual fidelity. These games do not faithfully replicate the appearance or behavior of objects warfighters must learn to recognize, and they do not provide a natural visual perspective on them. Physical fidelity of controls is also typically quite low (a gamepad may substitute for a steering wheel), and physical feedback (the impact of a bullet) is, of course, entirely lacking ingames. Nevertheless, fidelity is usually sufficient to recreate the key situations required for exercising more cognitive competencies, such as decision making, as well as team coordination.

remainder of this article, we describe this research and our findings concerning the functions of training systems that games fulfill, and those that must be developed to make games into training systems.

GORMAN'S GAMBIT

Gen. Paul Gorman (U.S. Army, Ret'd.) has asserted that commercial games can be used to train teamwork skills, regardless of whether the games depict modern or historic warfare (Gorman, July 2003). In a recent study conducted for DARPA and the Office of Naval Research, Aptima and prime contractor BBN tested this thesis (Weil, Hussain, Diedrich, Ferguson, & MacMillan, 2004). The findings from this study illustrate the potential of massively multiplayer games (MMPGs) for training, as well as their limitations.

Design: The research team evaluated several MMPGs to identify one that (1) simulated ancient or fantasy warfare, (2) supported as many as 40 players who were (3) differentiated by capabilities, (4) constrained their communications to a modern military hierarchy, (5) allowed players to affect each others' missions, and (6) enabled trainer/controllers to observe game play. Neverwinter Nights™ by BioWare Corp. was selected as the MMPG that best satisfied these requirements [2] (see Figure 2). The game was augmented with a Voice Over IP (VOIP) network.

Forty U.S. Army Infantry soldiers participated in the study of teamwork within Neverwinter Nights. The soldiers were between 19 and 33 years of age (M= 23.6 years) with 1.5 to 174 months of military experience (M= 51.2 months). Participants ranged from E-2 (Private) to O-1 (Second Lieutenant) (M= E-5, Sergeant). The participants averaged 51.2 months of military service, and 25 of the 40 had been deployed within the last year. They averaged 3.7 hours per week of computer use in the preceding year, and 4 hours of game-playing weekly on PCs or console systems (e.g., Playstation, Nintendo).

[2] Other games evaluated in this study were Ultima Online, trademarked by Electronic Arts, Inc.; Everquest, trademarked by Sony Computer Entertainment America Inc.; Dark Age of Camelot, trademarked by Mythic Entertainment, Inc.; World War II Online, trademarked by Playnet Incorporated; Battlefield, trademarked by Electronic Arts, Inc.; Final Fantasy, trademarked by Square Enix Co., Ltd; Star Wars Galaxies, trademarked by Lucasfilm Entertainment Company Ltd.; PlanetSide, trademarked by Sony Online Entertainment, Inc.; and Unreal Tournament, trademarked by Epic Games, Inc. Neverwinter Nights is a trademark of Wizards of the Coast, Inc.

In this study, the soldiers played on two competing teams of 20 members each, organized hierarchically as one platoon leader, and three squads each consisting of a squad leader and five or six specialized soldiers.

Figure 2: Opponents battle in Neverwinter Nights™ (Bioware Corp.)

The research team configured the game to require and afford a set of teamwork behaviors that are well established in the research literature: coordination through planning, leadership, monitoring, backup, communications push and pull, team orientation, and adaptation (Cannon-Bowers, et al., 1995; Freeman, et al., 2003; Sims, Salas, & Burke, 2004; Smith-Jentsch, et al., 1998). In particular, the authors systematically varied the capabilities of avatars. Each squad consisted of an archer, medic, artillery specialist, scout, and one or two strongmen, in addition to the squad leader. Tasks in the simulation space were designed to require coordinated use of their specialized skills. The VOIP communications network was configured to constrain communication realistically, allowing talk only between leaders at the two echelons, and within squads.

On the first day of the study, participants engaged in six hours of training to learn Neverwinter Nights and practice coordinating as a team. On the second day, participants executed three games of capture the flag. Measures of teamwork process and outcomes were taken by observers. Participants' observations and comments were also gathered in debriefings and in a post-study survey. This measurement strategy captured a convenient sample of teamwork behaviors (but did not systematically measure all instances of teamwork behavior).

Findings: The findings from this study illustrate the potential role of online games in training systems.

This carefully designed instance of a medieval military MMPG elicited teamwork behaviors in all of the targeted categories. The game provided practice in teamwork skills ($T_{simulator} \rightarrow H_{trainee}$) (see Figure 3). Most trainees (80%) stated this in their assertions that the game bore a functional relationship to modern military practice, and in particular that the communications system replicated current, hierarchical networks.

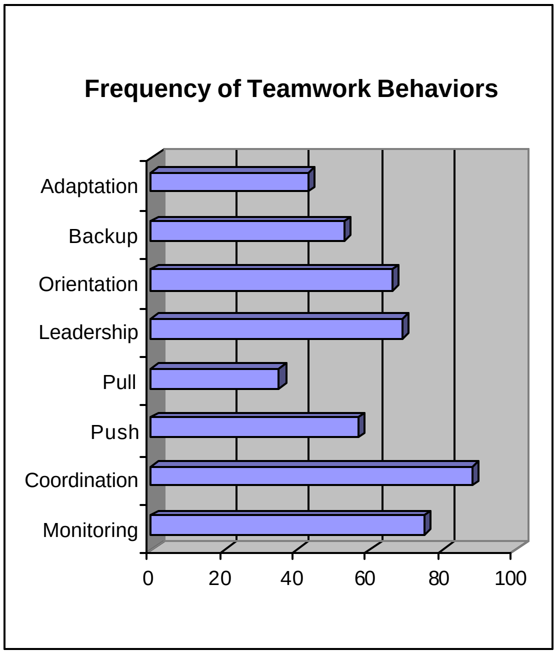

Figure 3: Gameplay in Neverwinter Nights™ elicited all of the targeted teamwork behaviors.

Games are often recommended for training because they are engaging. Ironically, participants in this study found this game only moderately exciting and interesting (rating it 4 on average on a scale of 0 to 7), and not at all stressful (rating: 1.7). There are several possible explanations for this finding. Trainees may not have reached the level of expertise with the interface or team strategies to fully enjoy and engage in the game. Alternatively, a more intensive scenario may have been needed, or a more rapidly paced study. Finally, it may be that some trainees were simply disinclined to engage with the online game. In the commercial environment, less interested users simply stop playing games. In training, all trainees *must* play; those disinclined to play may learn less. This aptitude (or attitude) x treatment interaction deserves further research.

Some functions of Neverwinter Nights proved particularly useful for training.

The ability to define avatars and the environment was critical to this study. Both were defined to support and require teamwork between participants, and in particular, teamwork related to several specific and empirically validated skills. Avatar and environment authoring is supported by some gaming systems. However, no authoring system to our knowledge constrains design so that it is relevant to training objectives. ($H_{trainer} \rightarrow T_{authoring}$). This is an open area for research and development.

AI entities in the game provided useful introductory training concerning NWN's capabilities and user interface ($T_{simulator} \rightarrow H_{trainee}$). However, it was necessary to supplement this with training ($H_{trainer} \rightarrow H_{trainee}$) and practice ($T_{simulator} \rightarrow H_{trainee}$) focused on avatar capabilities and teamwork prior to the start of the game.

The researchers were able to monitor trainee teamwork by using an observer function (called a Dungeon Master) within the virtual environment. This allowed them to survey the field of play unobserved by the trainees ($H_{observer} \rightarrow H_{trainee}$). In addition, researcher / controllers were able to manipulate the gamespace and avatars during play, as was the case when they "lifted" some avatars out of a space that trapped them artificially and unrealistically ($H_{controller} \rightarrow T_{simulation}$). Between games, the researchers used the rapid authoring tools of Neverwinter Nights to refine the capabilities of avatars (e.g., making platoon leaders more powerful and speedy) and the characteristics of some objects in the environment ($H_{controller} \rightarrow T_{simulation}$).

This game, like most, provided no functions for measuring human behaviors, such as teamwork ($H_{trainee} \rightarrow T_{measurement}$). Thus, observers manually rated performance, and participants' subjective ratings were recorded. This strategy is often necessary; always costly; and produces data that are incomplete, poorly sampled, and sometimes inaccurate due to the perceptual limitations of observers a host of nearly inescapable cognitive biases such as primacy, recency, anchoring (Kahneman, Slovic, and Tversky, 1982).

In addition, all debriefing was conducted by participants with no support from the gaming technology in the form of scenario replays or performance assessments. Many military trainers prefer this debriefing strategy because it engages trainees in critiques, diagnosis, and argument – all useful skills for learning and mission execution. However, some research is underway concerning the potential for automatically retrieving critical scenario events as an aid to debriefers (see T-CAST, below; $T_{feedback} \rightarrow H_{trainer}$), and automatically presenting assessments of performance for solo trainees (see STRATA, below; $T_{feedback} \rightarrow H_{trainee}$).

In sum, Neverwinter Nights provided a productive team practice environment and certain authoring and observer/controller tools. However, it was necessary to manually constrain authoring to focus on teamwork training objectives, and to supplement this game with human measurement and feedback processes.

T-CAST: TEAM COACHING ASSISTANT FOR SIMULATION-BASED TRAINING

In an ongoing effort for the U.S. Army Research, Development and Engineering Command – Simulation and Training Technology Center (RDECOM-STTC), Aptima and BBN Technologies are developing T-CAST (Team Coaching Assistant for Simulation-Based Training). T-CAST is an intelligent subsystem that augments the teamwork training capabilities of a MMPG: the Asymmetric Warfare - Virtual Training Technology (AW-VTT, see Figure 4) under development by Forterra Systems and RDECOM-STTC.

Figure 4: Forterra Systems' Asymmetric Warfare – Team Training Technology (AW-VTT) is a MMPG being developed for asymmetric warfare training.

AW-VTT is a prime example of a commercial MMPG technology that is being modified for use as a military training tool. Like Neverwinter Nights™, AW-VTT presents an environment in which human trainees can control avatars that execute military missions in a visually rich game space; however, AW-VTT simulates real-world operational environments (e.g., a modern urban landscape in the Middle East) and characteristic US forces and capabilities. AW-VTT is specifically designed to recreate conditions associated with asymmetric warfare. (The asymmetry reflects the discrepancy in strength between friendly and hostile forces. Such discrepancies encourage the use of non-traditional tactics by opposing forces and require similarly non-traditional tactics from friendly troops, such as specialized crowd control and building search procedures during military operations within urbanized terrain, or MOUT).

The function of T-CAST will be to recognize emergent teamwork events within AW-VTT and to index these for rapid retrieval and review during debriefing. A pilot study explored a method for developing basic training and assessment for a MOUT scenario simulated within AW-VTT.

Design: The pilot study consisted of three phases: scenario design, scenario execution, and development and evaluation of behavior models.

In the design phase, the team mapped extent teamwork theory (Sims, Salas, & Burke, 2004) to basic MOUT tactics, techniques, and procedures (TTPs) for conducting a street patrol with a squad of dismounted infantrymen. The researchers then developed a simple scenario that exercised these TTPs ($H_{trainer} \rightarrow T_{authoring}$) by (1) crafting a basic story with support from an SME, (2) fleshing out details in accordance with Army doctrine, and (3) modifying these details to align with the capabilities of the AW-VTT simulator. (For example, certain standard hand signals were modified to accommodate the gestures currently available to avatars). The team additionally identified potential errors that could occur during the scenario, either as omissions of described behaviors or as incorrect deviations from the prescribed activities. Several of these errors were incorporated within the scenario script.

During scenario execution, six participants played out the scenario in AW-VTT, maneuvering their avatars along streets, across intersections, and around doorways and other potential threat locations in accordance with Army doctrine ($H_{trainee} \rightarrow T_{simulator}$) (see Figure 5). Participants replicated both correct and incorrect behaviors defined within the scenario script.

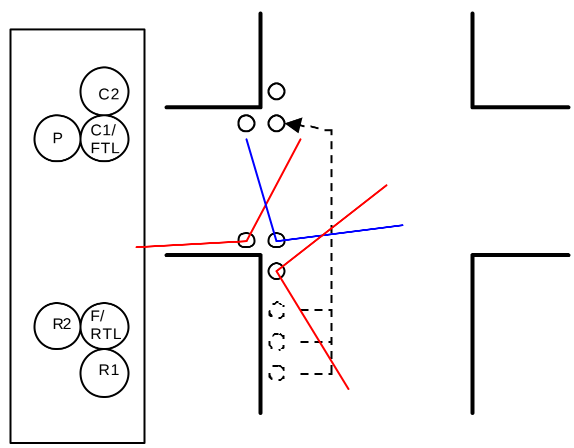

Figure 5: Crossing an intersection, two fire teams cover and bound around one another. T-CAST will recognize the execution of maneuvers such as this.

In the modelling phase, we prototyped human behavioral models that captured the activities described in the scenario script. We first developed a set of instructions for each player that explicitly defined the doctrinal behaviors that players were expected to demonstrate through control of their avatars within the scenario, as well as the conditions under which these behaviors were required. We encoded these instructions within finite state network models and validated these models by providing them with artificial simulator data and evaluating their responses to ensure completeness (all behaviors were accounted for) and consistency (there were no contradictions or mismatched dependencies within or across player models).

Findings:

This exercise demonstrated a principled method for developing scenarios for use in MMPG-based military teamwork training. AW-VTT proved to be an effective platform on which to deliver this training, simulating many of the behaviors from which critical teamwork competencies and specific knowledge and skill deficits implicated in infantry TTPs could be assessed. In subsequent review of the simulation video, we verified that all correct and incorrect behaviors described in the scenario scripts were observable and classifiable.

The effort also resulted in the development of human behavioral models that illustrate an elementary capability (now being refined and implemented) to encode typical tactical problems and solutions. This capability will ultimately allow the T-CAST pattern-matching subsystem to recognize and record instances of key operational situations occurring within the simulation environment ($T_{simulation} \rightarrow T_{measurement}$), as well as trainee responses to these situations (as demonstrated through the actions of their avatars – $H_{trainee} \rightarrow T_{measurement}$).

We concluded from this brief feasibility study that the AW-VTT MMPG supports presentation of scenarios that tap teamwork competencies. However, a principled process for authoring scenarios will be required to ensure that practice addresses teamwork skills. In addition, we concluded that it is possible to model and automatically recognize both the emergent situations in which teamwork skills should be exercised and the exercise of those skills. These functions can facilitate performance assessment and evaluation, critical components required for transforming games into training.

STRATA

U.S. Navy aviators practice coordinating with other air wing elements during exercises at the Naval Strike and Air Warfare Center (NSAWC) air wing training detachment. However, this coordination training is expensive, logistically complex, and, thus, rare. Between the infrequent exercises, trained skills degrade quickly (Arthur, Bennett, Stanush, & McNelly, 1998).

As part of the DARWARS program (see above), several organizations are collaborating to augment a high fidelity, laptop flight simulator (Airbook) with a variety of training functions. The product, STRATA, will provide Synthetic Teammates for Realtime Anywhere Training and Assessment (Bell, Johnston, Freeman, and Rody, 2004). The project team is led by CHI Systems and includes Aptima, Visual Training Solutions Group, and NAVAIR / Orlando.

Airbook (by VTSG) is a high-fidelity simulation of the displays, flight controls, weapons, sensors, and flight characteristics of an F/A-18 fighter jet. It runs on a laptop augmented with two, high-quality flight controls. An HLA interface (by MÄK Technologies) provides interoperability between the simulation and other components of the STRATA training system. The most critical of these are the following.

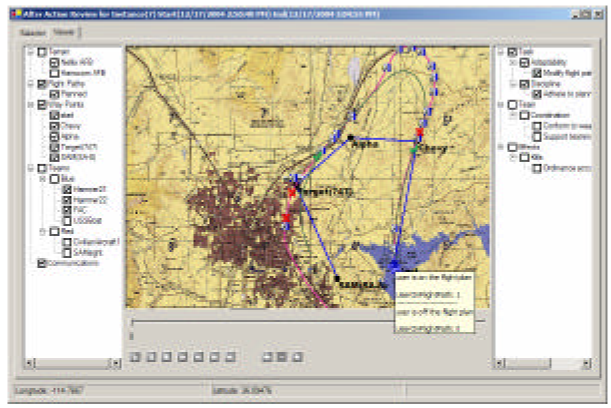

Figure 6: The STRATA AAR system provides user and automated control of map overlays (left pane), a tactical replay view with rollover details concerning entities and assessed actions (center), and control (right) over which assessments are displayed, by training objective.

Behavioral models (created by CHI Systems) serve as intelligent, interactive teammates in STRATA's close air support scenarios (CAS). The models can fly an Airbook F/A-18 as wingman or flight lead ($T_{ai} \rightarrow T_{simulator}$), serve as Forward Air Controller, and communicate verbally with the human user in either role ($T_{ai} \leftrightarrow H_{trainee}$).

The STRATA training configuration system takes data concerning training mission (e.g., CAS), user role (e.g., flight lead), training objectives (e.g., situational awareness, adherence to planned route, response to threats, communications), and other data to automatically populate a brief, select and configure performance measures ($T_{LMS} \rightarrow T_{measurement}$), and select a debriefing format.

The performance measurement system leverages verbal communications data from the behavioral models, and event data from the simulation to measure and assess performance on each active training objective. These assessments are presented in a debriefing system that supports the traditional review by replay. It also supports review by training objective, in which the user selects training objectives of interest, the After Action Review (AAR) marks the location of relevant assessments using icons on a tactical map, and the user replays the action and communications at those points to diagnose failures and successes (see Figure 6; $H_{trainee}$ → $T_{feedback}$). Future plans call for an automated debriefing presentation in addition to the manual modes just described. ($T_{feedback}$ → $H_{trainee}$).

In sum, STRATA augments a laptop simulation with HLA, AI, briefing, measurement, and feedback technology. All of these are important components of a training system.

CONCLUSION

In the studies and development efforts described above, Aptima and its partners are attempting to define the role of online games in training, augment the capability of those games, and test the effects on practice and, eventually, on learning.

There are several ways to satisfy some of these requirements: technology external to games (such as voice communication systems), technology integrated into games (such as measures), or processes (e.g., methods for designing avatars, environments, and events so that they are relevant to training objectives). Future research and development should evaluate these and other strategies, and determine how to efficiently augment games to satisfy the requirements of training.

Error! Not a valid bookmark self-reference. summarizes functions that concern us most because they are essential to a robust training system but often lacking in online games.

ACKNOWLEDGEMENTS

This work was supported by DARPA, the Office of Naval Research, and U.S. Army RDECOM-STTC. The opinions expressed here are those of the authors, and not necessarily those of the U.S. Department of Defense.

Table 1: Functions required to create training systems using online games. (→ denotes control flow or information flow).

Relation	Required training function
$H_{trainer}$ → $T_{authoring}$	Guide the design of environments, avatars, and events so that they are relevant to the training audience & training objectives
T_{LMS} → $T_{measurement}$	Support rapid game configuration / customization for the trainees and training objectives at hand
$T_{simulator}$ / $H_{trainer}$ → $H_{trainee}$	Train the use of game controls, team roles, and teamwork behaviours
$H_{observer}$ / $T_{simulation}$ → $H_{trainee}$	Enable observers to monitor gameplay
$T_{simulator}$ ←→ $H_{trainee}$	Provide trainees with the perceptual cues and action controls required to train targeted skills
T_{ai} → $T_{simulator}$	Provide intelligent teammates and opponents required to create a realistic mission environment
T_{ai} ←→ $H_{trainee}$	Enable intelligent entities to communicate and interact with trainees
$H_{controller}$ → $T_{simulation}$	Enable controllers to change the environment & events during game play
$T_{simulation}$ → $T_{measurement}$	Recognize emergent opportunities to test & measure targeted skills
$H_{trainee}$ → $T_{measurement}$	Measure the behaviors and effects of individuals and teams
$H_{trainer}$ / $H_{trainee}$ ←→ $T_{feedback}$	Enable trainers & trainees to retrieve specific assessments and replays in debriefs
$T_{feedback}$ → $H_{trainee}$	Automatically, intelligently display assessments and replays to trainees

REFERENCES

Arthur Jr., W., Bennett Jr., W., Stanush, P. L., and McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance*, 11, 57– 101.

Bell, B., Johnston, J., Freeman, J., & Rody F. (2004). STRATA: DARWARS for Deployable, On-Demand Aircrew Training. In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, December, 2004.

Cannon-Bowers, J.A., Tannenbaum, S.I., Salas, E., & Volpe, C.E. (1995). Defining team competencies and establishing team training requirements. In R. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 330-380). San Francisco: Jossey-Bass.

Chatham, R.E., and Braddock, J. (2001). Report of the Defense Science Board Task Force on Training Superiority and Training Surprise (Washington, DC: Office of the Undersecretary of Defense for Acquisition, Technology, and Logistics, 2001), 5, http://www.acq.osd.mil/dsb/ trainingsuperiority.pdf

Chinese online game sales nearly doubled. (16 January 2004). ITFacts. On line at: http://www.itfacts.biz/index.php?id=P562

Freeman, J., Diedrich, F. J., Haimson, C., Diller, D. E., & Roberts, B. (2003). Behavioral representations for training tactical communication skills. *Proceedings of the 12th Conference on Behavior Representation in Modeling and Simulation*, Scottsdale, AZ.

Gorman, Paul. (July 2003). Comments at DARWARS program meeting. Washington, D.C.

Kahneman, D., Slovic, P. & Tversky, A. (Eds.). (1982). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge University Press

Online game sales surpass $1 billion in U.S. (24 March 2004). Mercury News. On line at: http://www.siliconvalley.com/mld/siliconvalley/8263163.htm

Sims, D.E., Salas, E., & Burke, C.S. (2004). Is there a "Big Five" in teamwork? 19th Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

Sims, D.E., Salas, E., & Burke, C.S. (2004). Is there a “Big Five” in teamwork? 19th Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

Smith-Jentsch, K.A., Zeisig, R.L., Acton, B., & McPherson, J.A. (1998). Team dimensional training: A strategy for guided team self-correction. In J.A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 271-295). Washington, DC: American Psychological Association.

Weil, S.A., Hussain, T.S., Diedrich, F.J., Ferguson, W., & MacMillan, J. (2004). Assessing Distributed Team Performance in DARWARS Training: Challenges and Methods. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference*, Orlando, FL.