Feedback Mechanisms in Automated Emergency Management Training

Matthew Guardascione
Monmouth University
matthew.guardascione@gmail.com

Allen E. Milewski
Monmouth University
allen.milewski@monmouth.edu

ABSTRACT

This study explored automated training for emergency managers and the effects of feedback on performance. A prototype emergency management training application was built to allow the usage of either immediate feedback or delayed, "hotwash" feedback. Users were split into two groups and asked to carry out two emergency management scenarios using one of the feedback mechanisms, and the difference in scores between each feedback type were analyzed There was a general increase in performance across sessions. Further, the improvements in scores between each feedback type showed that users performed significantly better when using the hotwash feedback mechanism compared with the immediate feedback mechanism. In contrast to the performance data, preference data showed no overall differences between the two procedures, although each user had a strong preference for one or the other feedback mechanism. The implications for the design of training systems offering both procedures are discussed.

Keywords

Emergency management, training, immediate feedback, delayed hotwash feedback

INTRODUCTION

Training emergency workers and first responders poses a complex set of challenges. First, because a large proportion of these workers are volunteers, (Pinkowski, 2008) turnover rates can be moderate to high and individuals' involvement can be sporadic. Therefore, training has to be provided on a frequent and convenient schedule. Second, training must take place at several cognitive levels, covering a wide range of basic skills, broad procedural knowledge and flexible problem solving and improvisation strategies (Mendonca, Beroggi, van Gent and Wallace, 2006). For example, while it is necessary to perform many specific, fundamental procedures in emergency situations, key training experiences must also provide extensive practice in deciding when to use each procedure. Finally, since many emergency situations are inherently dangerous, new trainees must acquire experiences which involve the practical application of their knowledge and procedures before entering into a real crisis. Based on these challenges, one might expect automated training of emergency workers to be common. However, despite advancements in technology over the past few decades, most emergency management training is still done manually, with very little automation or use of computer systems. The training framework provided by the U.S. Department of Homeland Security outlines nine different types of acceptable emergency training (Renner, 2001; U.S. Department of Homeland Security, 2007). Of these, the most common types are the table-top exercise and the full-scale walkthrough. In table-top exercises, a human facilitator with high expertise in all aspects of incident response leads a group of trainees through a semi-structured, verbal or written emergency scenario. The facilitator poses questions, notes tradeoffs, decision-points and potential strategies and provides helpful tips. The trainees are typically co-located and are often given specific roles to perform within the response team as they decide upon appropriate actions and convey them to the facilitator. In full-scale exercises, trainees are immersed in a quasi-realistic physical setting and experience a predetermined set of events that mimic a real emergency, often including unexpected "insert" events. Trainees are expected to interact with others and respond according to standard procedures but also in such a way as to minimize casualties, injuries, material loss, etc. In both table-top and full-scale exercises, a critical element is a post-exercise "hotwash" session where expert facilitators review trainees' activities, evaluate their performance and make suggestions for improvement. Such after-action analyses have been shown to be an effective pedagogical tool (Nilson, 2009; Schaafstal, Johnston, and Oser, 2001). In sum, these key training procedures are human-intensive and while there have been some automated-training products introduced (e.g. EMST; Pigora, Tamash and Baxter, 2006), the majority of emergency management training remains manual.

Reviewing Statement: This paper has been fully double blind peer reviewed.

PEDAGOGICAL AGENTS AND SERIOUS GAMES

While computer-based training is not the norm in the emergency management field, various forms of automated training have computer been found effective in other domains and these systems can be roughly divided into two types: pedagogical agents and serious games.

A pedagogical agent is an electronic learning tool which can help train, negotiate, and mediate students while they attempt to solve training scenarios. They help individuals learn material, providing additional information to students in order to help give them the best possible education in a particular field- specifically by providing hints, additional factual material and alternative problem-solving strategies (Atkinson, 2002). Teachers have been using pedagogical agents for years in order to incorporate computer technology in the classroom to better help students understand important fundamental material (Kim, 2005).

"Serious games" are videogames which are developed for non-entertainment purposes (Moon, 2006; Wideman, Owston, Brown, Kushniruk, Ho, and Pitts, 2007; Toups, Kerne and Hamilton, 2009). A serious game provides a realistic training scenario but in a comfortable environment that may not be reachable in real time exercises and training (Toups and Kerne, 2009). The game can incorporate multi-player aspects which would allow individuals to have the same group experience that they would in a live full scale walkthrough (The Serious Game Initiative, 2008). Serious games can include pedagogical agents and often utilize a videogame graphical simulator to improve knowledge, group performance, and expertise in experience based training (Kim, 2005).

Research has suggested that videogames can provide effective learning environments. Gee (2003) has summarized at least 36 pedagogical principles common in videogames- even when used for entertainment. Among the principles Gee lists are the semiotic principle, where users learn interrelations across multiple sign systems (images, symbols, actions, etc), the psychosocial moratorium principle, where learners can take risks with no real world consequences, and the incremental principle, where learning starts generalized and gradually increases in complexity.

One tremendously useful capability of serious games is the ability to properly execute full scale walkthroughs without having to consume real time, money, or manpower. The concept of teaching individuals by networking in a massively multiplayer simulator has shown promise. Steinkuehler (2004) reports an experiment with a massively multiplayer online game to determine what can be taught to individuals. The results of the experiment showed that many of Gee's principles, including his semiotic domain and layered training principles, occur in the online game.

Training with Serious Games has promising applicability to the emergency management domain (Turoff, Chumer, Hiltz, Hendela, Konopka and Yao, 2006). With such games, trainees are able to interact with trainees in their field as well as other fields to experience a wide range of effective simulations. In addition to the realism of this, trainers are able to pause during the simulation to educate, quiz, and change the scenario to modify its difficulty for students to learn even better than the full scale scenarios. By doing this, the shortcomings of the full scale walkthrough (time, money, non-pause-ability, manpower, etc.) are remedied through the videogame.

Serious games can be effective teaching tools in both entertainment and educational environments (The Serious Game Initiative, 2008; Greitzer, 2007). They employ many sound pedagogical principles, many of which stress the importance of feedback as a critical factor. For example, several of Gee's principles are based on the educational importance that feedback can play in videogames. His Active, Critical Learning Principle, for example, shows the importance of the learning actively behaving and then observing the consequences of that behavior. His Discovery Principle and Probing principle suggests that this action/feedback cycle is an important way to learn by experimenting. Finally, the Amplification of Output principle claims that an instructive game is one that gives rich and ample feedback every time the learner makes any action. Moreover, this feedback has few real world consequences, thus educational risk-taking is encouraged (Gee, 2003).

Typically, videogames make use of the concept of "leveling-up" which gives the player nearly immediate feedback on their performance in the game, as well as acknowledgement of their accomplishments in the game. Further, immediate feedback adds the important feature of keeping players enticed with the game and wanting to progress further. Greitzer (2007) believes that, "this validation of improvement provides reinforcement for the learner. Game players love the feelings of 'getting better' at something – especially achieving mastery over something difficult and complex" (Greitzer, 2007). By adding leveling up to a video game, players are given positive feedback of their immediate actions (in the form of levels, points, progression into the game, etc.), which in turn entice them to continue playing the game. If applied to a serious game, users can be given immediate feedback of their actions in the game to allow them to learn from their mistakes, and will also give a player a desire to play the game further. Note that the practice of providing near-immediate feedback in videogames is in contrast to the form of feedback provided by the most common forms of emergency management. In this context, it is most common to rely upon "hotwash" feedback procedures where trainees receive comments on their performance after the training exercise is over. Hotwash sessions are typically held anywhere from immediately after the end of an exercise up to a few weeks afterwards.

Emergency management training can likely benefit from the utilization of serious games, but before extensive implementation of serious emergency management training games is undertaken, more research is needed to gain a detailed understanding of how best to implement various serious game features . The present study has focused on one such feature: feedback and its role in training. This study explores the effects of different forms of feedback on trainees' performance and preferences – specifically immediate and delayed, "hotwash" feedback - as they receive automated emergency management training.

METHOD

To provide a flexible platform with which to present interactive training scenarios to test trainees, a basic prototype training system was implemented which is shown in Figure 1.

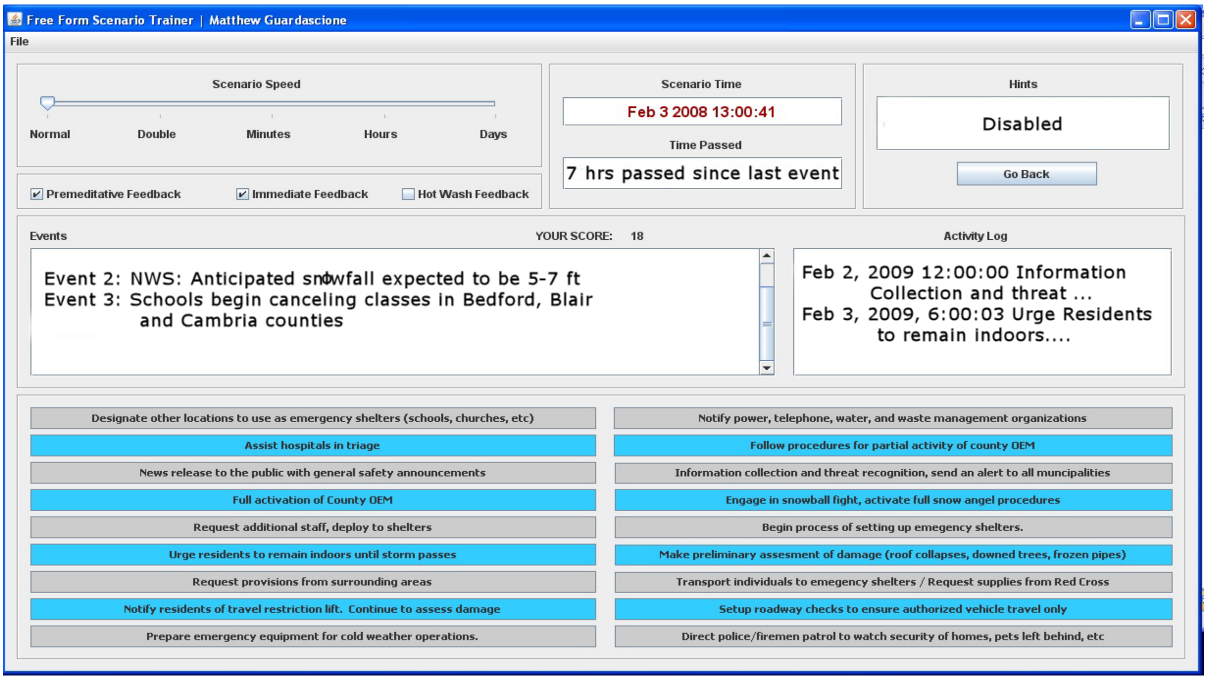

Figure 1. Basic Interactive Training System

The prototype training system provided a simulated emergency management environment as well as a measure of correctness regarding trainees' behaviors. Text descriptions of successive scenario events were displayed in the events window. Their speed of presentation could be controlled by the trainee with the "scenario speed" slider by increasing or decreasing the rate of scenario time passage. Every effort was made in the design of the prototype interactive trainer to maintain a realistically continuous flow of events and to avoid revealing when an action might be expected with the "multiple-choice test" flavor of many computer-assisted learning environments. Users could respond in a "free-play" fashion (Pigora, et al., 2006) - at any point in time- whether or not there was a new event being displayed. Alternatively, they could decide to do nothing in response to an event. The reason for this design decision was to implement a system that could ultimately be compatible with a much more sophisticated, simulation-based trainer being built in our laboratory. This trainer (The All Hazards Exercise Training Tool- AHETT) is a networked, multi-agent, artificial intelligence simulation environment that monitors trainees' activities as a scenario runs its natural course. In AHETT, activities are expected to be realistic ones such as telephone calls and EMS commands. For our simpler prototype training system, this design decision necessitated that all possible responses for the entire scenario be displayed to users at the same time along with some activities that would never be used. All these responses were displayed in a block of response buttons that were available throughout the exercise. These response buttons fill the bottom half of the display, and while not readable in Figure 1, display such actions as "Request additional staff, deploy to shelters" or "Assist hospitals in triage". Finally, an activity Log displayed for the trainee his past actions, and feedback was provided for all responses by updating a cumulative numerical Score. This score was calculated by using a numerical system based on the appropriateness of an answer at a certain point in time. Specifically, prior to the study, an experienced emergency management expert analyzed each scenario event, determined the most important action to be taken during it and assigned a score of 10 to that event action pair. If the action was somewhat appropriate during other events, a score of 9 or 3 might also be assigned to that pairing. During training, trainees received a score of 0-10 for each event depending on the action taken. A hotwash user's score would be determined by the accumulation of these points throughout the training, while an immediate feedback user's score was determined by taking the score of the first answer chosen so as to remain consistent with the scoring rubric used for the hotwash users. Other features shown in Figure 1, such as Hints and Feedback-type choice were not used in the present study.

The present study employed two feedback procedures:

1. Immediate feedback –this mechanism allowed trainees to know immediately after input the correctness of their response. A very important dynamic to this type of feedback, however, is to encourage trainees to learn from receiving this type of feedback. In order to accomplish this, the trainer allowed a user to select an answer, then gave them a score based on the correctness of their answer, and allowed them the option of going back, resetting time, and choosing another answer. Therefore, the immediate feedback mechanism used required users to "feel out" the best answer, rather than expect that answer to be simply handed over by the system.
2. Hotwash feedback – this "delayed" feedback gave trainees a complete analysis of their performance during the training – but provided only after the entire scenario was completed. This feedback included their responses, their corrected responses, and text descriptions of relevant hints to study in order to increase their performance when running through the exercise again. Unlike the immediate feedback, users were provided with what the system considered the "perfect" response set so they could improve on their performance on subsequent runs through the scenario. An example of hotwash feedback for one event is shown in Figure 2.

Trainees had no prior experience in emergency management. Two groups of ten trainees received two sessions each. The first session was used to vary feedback type while the second session was used as a measure of training. One group of ten trainees received immediate feedback on their first session; the other group of ten trainees received delayed, hotwash feedback on their first session. For both groups, trainees received no feedback at all during the second session. Two separate scenarios were created for the purposes of testing: a Nor'easter scenario, and a Blizzard scenario. Both scenarios reflected events that would occur within typical emergencies and were approximately the same length. The order of scenario was counterbalanced across groups. Different scenarios were used during the first and second session in an attempt to avoid measuring simple memorization of correct responses and instead to explore more general understanding of emergency management procedures.

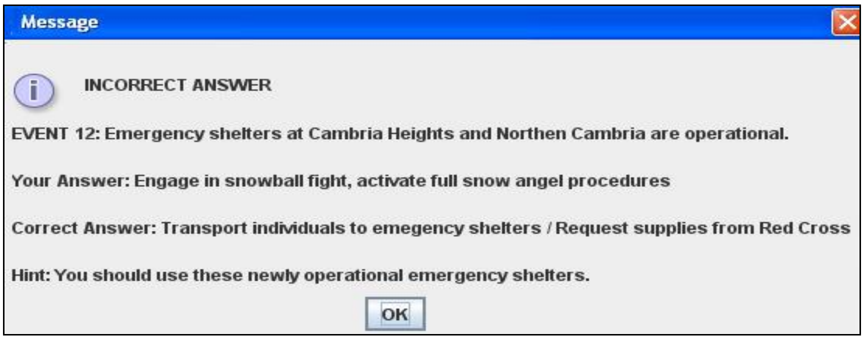

Figure 2. Hotwash Feedback for One Event

The experiment was then sent out to the separate groups and each was given about two weeks to take the test. The users returned the scores of the two scenarios, as well as a qualitative user feedback questionnaire to obtain their opinion of the exercise overall.

RESULTS

The results for Session 1 and 2 for all trainees are shown in Figure 3. The performance data across sessions were analyzed using a repeated measures analysis of variance which suggested a marginally significant improvement in performance from session 1 to session 2, ($F(1,36) = 3.80$, $p < .06$). While the interaction between sessions and scenario order did not reach significance, the increase between sessions was less for those trainees receiving the Blizzard scenario during the second session because, as is shown in Figure 4, that scenario was more difficult and resulted in lower performance no matter which session it was experienced.

In order to explore the effects of feedback type and scenario order on training performance during the two sessions, a multivariate analysis of variance (MANOVA) was performed. This analysis revealed a significant effect of feedback type on performance during the second training session ($F(1, 16) = 9.809$, $p < 01$. As shown in Figure 4, scores were higher for hotwash feedback groups than for immediate feedback groups. There was no such significant feedback effect, however for performance scores during the first training session ($F(1, 16) = 2.12$, $p = .116$), suggesting that feedback type differences were part of the learning taking place between the two sessions. In addition to the feedback effect, there was also an effect of scenario order on Session two performance with those trainees receiving the Nor'easter scenario after the Blizzard scenario performing better than those receiving the opposite order ($F(1, 16) = 6.434$, $p < .03$). This is shown in Figure 4. There were no significant interaction effects of feedback or order.

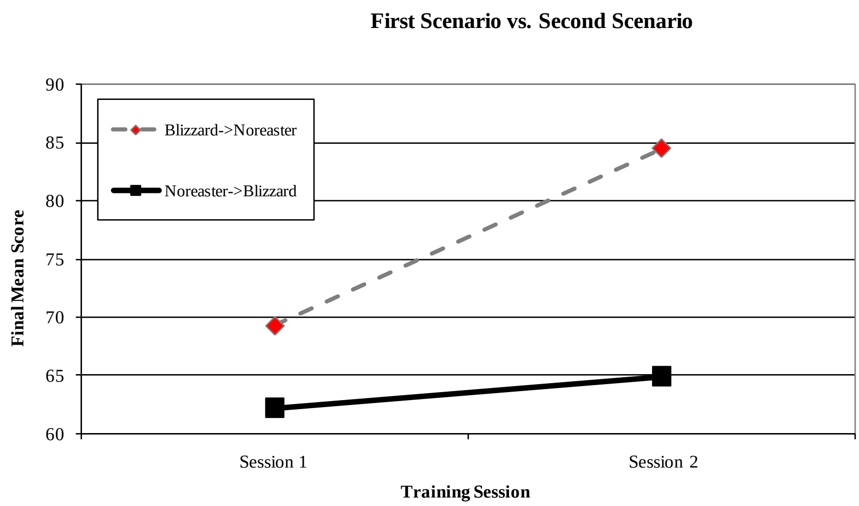

Figure 3. Mean Score by Training Session

In addition to performance data, qualitative preference data were collected with a questionnaire administered after the training sessions. The questionnaire showed a generally positive attitude towards the prototype trainer as a whole. However, on the topic of feedback, trainees tended to either strongly favor or strongly oppose the feedback type they experienced during the training sessions. An equal number of trainees expressed a preference for hotwash and immediate feedback. Interestingly, trainee's preferences were not a function of their own experience. Eleven out of the 20 trainees expressed negative attitudes toward the feedback mechanism they experienced or would have rather had the other group's feedback mechanism.

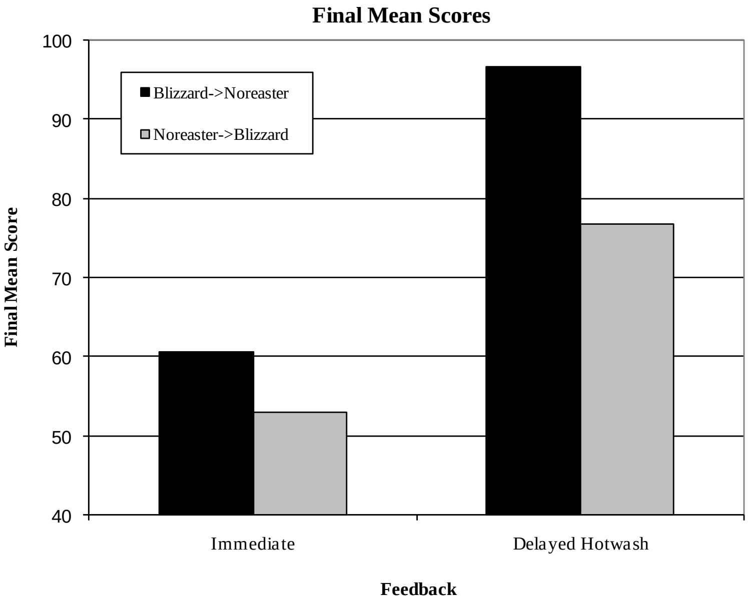

Figure 4. Performance Scores on Session 2 by Feedback Group

DISCUSSION

One of the intended goals of the experiment was to explore whether measureable learning could be demonstrated in the context of an automated Emergency Management training system. The increase in performance scores from Session 1 to Session 2 suggests that learning can occur when using an automated trainer, and is consistent with previous studies suggesting the potential for automated training (Turoff, et al., 2006) It can further be argued that trainees learned, at least to some extent, general principles about emergency management procedures and strategies rather than responding at a simplistic level to event conditions. This is argued since the improved performance appeared to generalize across differing scenarios across sessions. All of this was possible using an automated tool with no human intervention, suggesting that it may be possible for an automated Emergency Management trainer to someday remove the need for a facilitator. Of course, more research is needed to explore the generality of these findings and especially transfer of learning to performance in real world crises.

The other goal of the experiment was to explore feedback procedures and their effectiveness in training. The results of the experiment show that the delayed, hotwash feedback procedure resulted in better performance improvement compared with immediate feedback. This finding in the context of automated training is entirely compatible with the most common forms of feedback in current, non-automated emergency management training. On the other hand, these results are in contrast with our initial expectations based on analogies with serious games which often use relatively immediate feedback. There are several potential explanations. First, while the hotwash feedback provides both correct answers and also explanations that might facilitate a more complete understanding, the immediate feedback procedure only modifies and informs trainees of their correctness or incorrectness by modifying their numeric score. As one immediate feedback trainee notes:

"I found myself looking more for a +10 in my score rather than a correct answer, so I wound up retaining only a positive outcome rather than remembering what the correct answer was."

Second, previous research has shown similar results when immediate and delayed feedback are compared in an educational setting. Butler, Karpicke, and Roediger (2007) report that delayed feedback produced better long-term retention compared with immediate feedback. The study goes on to explain that, "Information is better retrained when learned through repeated presentations that are spaced (or distributed) as opposed to massed. After a correct response, delayed feedback would represent a spaced presentation of the information, whereas immediate feedback would represent a massed presentation." Similarly, Schooler and Anderson (1990), found that while their immediate feedback group finished the training material in 40% less time, the delayed feedback users retained knowledge of the correct answers better. The researchers suggest that, "immediate feedback competes for working memory resources, forcing out information necessary for operator compilation. In addition, more delayed feedback appears to foster the development of secondary skills such as error detection and self-correction, skills necessary for successful performance once feedback has been withdrawn". This suggests that new trainees may not be able to take advantage of immediate feedback because of information overflow. However, the advantage of delayed over immediate feedback is limited to certain contexts. Butler, Karpicke, and Roediger (2007) note that studies have shown users benefit from delayed feedback in laboratory settings, although benefits from immediate feedback can be seen in classroom settings. The material by which the evaluations are given, such as digitally, written, or orally, can also change the effectiveness of the feedback mechanisms.

These contextual effects may have been reflected in the preference data collected in the present study. Many trainees noted that taking the time after the training scenario to carefully study their actions and the correct actions were more beneficial than doing the explorative "answer-until-correct" immediate feedback procedure. Still, other trainees credited their success to a matter of preference; depending on how they tended to learn as an individual, one feedback procedure proved more useful than the other. The presence of such strong preferences and context effects suggest that future automated training systems used for Emergency Management should provide trainees with a choice of both types of feedback.

ACKNOWLEDGMENTS

This research was done under the auspices of the Monmouth University Rapid Response Institute (RRI) and was funded in part by the Department of Defense Contract No. W911SR-07-C-0074. We express thanks to Dr. W. Tepfenhart, Dr. B. Reagor and Professor R. Kelly.

REFERENCES

1. Atkinson, R.K. (2002) Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology*, 94, 416-427.
2. Butler, A.C.., Karpicke, J.D. and Roediger, H.L. (2007) The Effect of Type and Timing of Feedback on Learning from Multiple-Choice Tests. *Journal of Experimental Psychology: Applied*, 13, 273-281.
3. EMST, Emergency Management Staff Trainer , http://emst01.jsrts.org/emst/data/portal/emstBrochure.pdf
4. Gee, J. P. (2003) What videogames have to teach us about learning and literacy, Palgrave Macmillan., New York.
5. Greitzer, F., Kuchar, O., and Huston, K. (2007) Cognitive science implications for enhancing training effectiveness in a serious gaming context. *Journal of Educational Resources in Computing* 7, 2-16.
6. Kim, Y. (2005) Agent learning companions: Learners expectations of the desirable characteristics. *Proceedings of Agent-based Systems for Human Learning,* AAMAS Workshop.
7. Mendonca D., Beroggi G.E.G., van Gent D., Wallace W.A. (2006) Designing gaming simulations for the assessment of group decision support systems in emergency response , *Safety Science*, 44, 6, 523-535.
8. Moon, I (2006) Evolution of Player Skill in the America's Army Game. *Simulation.* 82, 11, 703-718.
9. Nilson, J. (2009) Using tabletop exercises to learn about crises management: empirical evidence. *International Journal of Emergency Management,* 6, 136-151.
10. Pigora, M.A., Tamash, T. Baxter, H.C. (2006) Training Novices and Experts: A common Assessment Mechanism for Knowledge, Skills and Abilities, *Proceedings of the Interservice/Industry Training, Simulation & Education Conference (I/IT SEC),*
11. Pinkowski, J. (2008) Homeland Security Handbook, CRC Press, Boca Raton, FL.
12. Renner, Sarah. (2001) Emergency Exercise and Training Technique, *Australian Journal of Emergency Management*, 16, 2, 26-41.
13. Schaafstal, A. M., Johnston, J. H. and Oser, R.L. (2001) Training teams for emergency management, *Computers in Human Behavior*, 17, 615-626.
14. Schooler, L. J., and Anderson, J. R. (1990) The Disruptive Potential of Immediate Feedback. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* , 702-708.
15. Steinkuehler, C. A. (2004) Learning in massively multiplayer online games. *Proceedings of the 6th international Conference on Learning Sciences* , Santa Monica, CA, 521-528.

16. The Serious Game Initiative (2008) , http://www.seriousgames.org/
17. Toups, Z. O. and Kerne, A. (2009) Making Invisible: Communication as Core Mechanic in Non-Mimetic Simulation Games. *The Future of Interactive Media: Workshop on Media Arts, Science, and Technology*.
18. Toups, Z.O., Kerne, A, and Hamilton, W. (2009) Game Design Principles for Engaging Cooperative Play: Core Mechanics and Interfaces for Non-Mimetic Simulation of Fire Emergency Response. *Proceedings of the ACM international conference on Supporting group work*, Sanibel Island, FL.
19. Turoff, M. Chumer, M., Hiltz, S. R., Hendela, A., Konopka, J. and Yao, X. (2006) Gaming Emergency Preparedness, *Proceedings of the 39th Annual Hawaii International Conference on System Sciences, 2, 38.*
20. U.S. Department of Homeland Security. *Homeland Security Exercise and Evaluation Program (HSEEP)*. Vol. 1. Feb 2007.
21. Wideman, H.H., Owston, R.D., Brown, C., Kushniruk, A., Ho, F. and Pitts, K.C. (2007), Unpacking the potential of educational gaming: A new tool for gaming Research, *Simulation & Gaming*. 38, 11, 10-30.