Running Head: Training Experienced Combat Identification Teams

Team Training Paradigm for better Combat Identification

Wayne Shebilske

Wright State University

Georgiy Levchuk, and Jared Freeman

Aptima, Inc.

Abstract

We developed and tested a paradigm to train experienced teams for combat identification in simulated Air Force Intelligence, Surveillance, and Reconnaissance (ISR) and Dynamic Targeting Cell (DTC) teams. Teams performed about 16 hours of simulated missions, which had many differences, such as different time sensitive targets, and many similarities, such as particular areas having a higher probability of specific enemy threats. The pattern of differences and similarities then changed in ways that occur in operational settings. We present evidence that such changes are challenging, that they can be learned, and that this learning should foster better combat identification.

Team Training techniques for better Combat Identification

Combat Identification in the Air Force Offensive Team entails, not only recognizing what an object of interest is, but also whether the object of interest is an expected or unexpected threat or opportunity that demands a response. Expected threats and opportunities as well as planned responses to them are included in an Air Tasking Order, which is updated every 8 to 12 hours. Unexpected threats and opportunities that demand immediate action are called time sensitive target (TSTs). Two parts of the Air Force Offensive Team, ISR and DTC, play important roles in responding to TSTs. After ISR nominates an object of interest as a potential TST, the DTC differentiates TSTs from other unexpected objects, designates an object of interest as a TST, and submits plans to prosecute all designated TSTs. The plans include the order for attacking TSTs and strike packages for the attacks. A Senior Offensive Duty Officer (SODO) and the Chief Combat Officer (CCO) evaluate whether the proposed plans achieves the goals of prosecuting the TST efficiently and effectively without jeopardizing the ATO. Accomplishing these goals requires the DTC to identify enemy targets as well as friendly assets that have a potential contribution to a strike package for a TST greater than the potential cost of the asset not being available to execute the ATO.

The present experiment is part of a more general investigation of human factors issues related to ISR and DTC teams who confront changes that are challenging with respect to their past experience. We observed, for example, human factors issues related to displays. On the one hand, computer scientists and engineers have succeeded in storing within computer systems all the information that the Air Force Office team needs to execute an ATO and to respond to TSTs. On the other hand, DTC teams often cannot find and use the information fast enough to prosecute TSTs in time. Our general investigation therefore includes efforts to develop better displays and performance support systems. Similarly, our present investigation of training experienced teams addresses only a subset of the human factors issues related to training in this context. For example, our general investigation also includes human factors issues related to training novices to become experts in these complex tasks. We will return to these issues in our discussion section where we will consider how our present analysis of human factors issues in experienced teams learning new experiences relates to a more general perspective.

Experienced Air Force Offensive teams train for new experiences related to their current enemy escalating or changing tactics or to confronting their current and/or new enemies in different contexts. Such training usually happens on the job. Training with realistic simulations could have at least three advantages over training on the job. First, training systems would be highly accessible to promote frequent practice. Second, practice scenarios in simulations could be systematically structured and focused on training objectives, to promote deliberate practice. Third, performance assessments could be relevant to training objectives, and systematically delivered to train and maintain competencies.

A large body of literature on which human factors psychology is founded supports the use of computer-based training simulations and provides valuable training principles (e.g. Piorolli and Anderson, 1985; Schmidt and Bjork, 1982). Much of this literature, however, concerns basic training on tasks that are well-defined (e.g. mathematics) and that are executed mainly by individuals. In contrast, Air Force Offensive Teams and other command and control operations involve human performance by teams in domains that are ill-defined. Furthermore, the training under consideration is not initial training of new operators, but advanced training of experienced operators learning new experiences. The historical focus of rigorous research on beginners learning well defined tasks made sense not only because it provided theoretical foundations for understanding learning and practical foundations for applying learning principles to training in operational settings, but also because research tools were not available to enable comparable research on experts learning ill-defined advanced skills. Until recently, research on such complex domains was limited to explorations of operational settings with little rigorous hypothesis testing.

Using modern simulation tools, we are attempting to extend to more complex domains the historically established path of using computer-based training systems for rigorous laboratory research aimed at building a foundation for supplementing on-the-job training with realistic simulation-based training systems. Specifically, we are using Aptima's Dynamic Distributed Decision making (DDD) synthetic task environment to simulate and explore complex domains, precisely manipulate and measure variables that will guide modeling of human information processing during complex activities, and rigorously test well founded hypotheses about complex processes.

Our manipulations, measures, and hypotheses relate to our initial goal of developing and testing a paradigm to train experienced teams for combat identification in simulated ISR and DTC teams. In a previous experiment, Shebilske et al. (2007) manipulated a factorial combination of two independent variables: a) Phase II vs. Phase III training and b) Pretest vs. Posttest before and after each phase. In Phase I (this was not included in the hypotheses) simulated ISR and DTC teams performing 50 hours of background training. In Phase II, the same teams performed simulated missions for 49 hours. In Phase III, the same teams performed 18 more hours of simulated missions that were more difficult because the enemy increased the number of TST and the number of threats protecting each TST. The dependent variable was expert ratings of the DTC coordinating a strike package plan. Two experts made reliable independent ratings on a six-point scale. These ratings were enabled by a precise playback tool in Aptima's DDD 4.0. The tool integrated, time stamped, and played back all stimulus events and all responses in realistic simulations of ISR and DTC operations.

If college students can learn realistic ISR/DTC simulations, if they can be challenged by enemy escalations, and if they can learn to over come this challenge, then the following hypotheses should be supported:

1. Posttest scores should be higher than the pretest scores in Phase II.
2. Pretest scores on Phase III, should be lower than posttest scores in Phase II.
3. Posttest scores should be higher than the pretest scores in Phase III.

Shebilske et al. (2007) supported hypotheses 2 and 3, but not Hypothesis 1. They concluded that the pretest in Phase II was higher than expected suggesting that too much background training had been given before the pretest. In addition to the quantitative results, a qualitative result was interesting. Scientists who had performed task analysis of DTC dynamics in operational settings, observed the simulated DTC near the end of Phase II. The scientists observed that the simulated DTC team (college students) performed DTC dynamics at about the level of an experienced operational DTC with a medium-high skill level.

The present experiment reduced the amount of background training and other training in an effort to support Hypothesis 1, and it retested Hypotheses 2 and 3 to determine whether they would be supported with reduced training.

Methods

Participants

The participants were 7 undergraduate college students (2 women and 5 men, mean age = 20 years), who were paid $7.25 per hour for 36 hours each. Their participation was part of their responsibilities as research assistants. Although the trainees were research assistants, they did not know the purpose of the experiment.

Materials

Office dividers separated seven work stations with four stations in one row and three in the other. Each station had an IBM compatible PC with a 17 in. monitor, a mouse and keyboard for inputs, and a headset linked with an Aardvark sound system audio net that enabled open and recorded communication within teams and isolation of sounds outside the team.

We employed Aptima's DDD synthetic task environment to simulate the ISR/DTC task. We chose this task because of its relevancy to modern Air Force operations, to combat identification, and to general online command and control operations.

The synthetic task environment was based on a task analysis that we conducted on operational DTC teams. The Air Force structures DTC operations as a kill chain, which includes multiple stages and multiple task objectives (TOs). Our three TOs were: TO1, detect and differentiate TSTs; TO2, prioritize the TSTs; and TO3, coordinate attack assets. They relate to complex teamwork occurring in three of the kill chain stages: Find, Fix, and Target. Operations occurring during the Find stage are limited to the ISR who initiates operations related to TO1. The ISR role was filled by one of our trainees. DTC operations in the Fix phase include complex communications among the four DTC team roles, all of which were filled by our trainees: DTC Chief, Ground Track Coordinator (GTC), Attack Coordinator (AC), and Target Duty Officer (TDO). This teamwork relates to TO1 and TO2. DTC operations in the Target phase also include complex communications among the four DTC team members. This teamwork relates to TO3. In many operational settings and in the DDD simulation, the proposed order of attacking TSTs and the proposed strike package for each attack is approved or disapproved first by the SODO and then by the CCO. These two officers also control the execution of accepted proposals. In the present experiment, expert confederates played the roles of CCO and SODO proficiently so that variability related to approvals and to executions was minimal.

The confederates and the trainee who played the role of the DTC Chief sat in one row separated by a barrier from a row of the other 4 trainees. During operational missions and during the DDD simulated missions, these offensive team members usually communicate by text messaging and occasionally communicate by voice. The background training for the present trainees included standard Air Force brevity procedures for both communication modes. For example, the trainees were taught that the standard text or verbal acknowledgment to a message is “copy.”

The DTC must operate in the face of uncertain information and within the constraints of an existing mission plan, the ATO, which changed for each mission in the present experiment. Missions were usually separated by about 24 hours, but occasionally separated by 8, 12, or .5 hours in order to accommodate individual schedule demands. The ATO includes guidance on which objects of interests should not be prosecuted, what kinds of TSTs might be encountered, and how they should be prioritized relative to the ATO and relative to each other.

Design

The within-participant, independent variables were Phase (II vs. III) x Test Type (pretest vs. posttest). Phase I was 16 hours of background training; Phase II was 11 hours of performing missions, Phase III was 9 hours of performing more difficult missions.

The planned comparisons were the pretest versus the posttest in Phase II, which reflected the change in performance during that phase; the posttest in Phase II versus the pretest in Phase III, which reflected the drop in performance during the initial exposure to the new experience; the pretest versus the posttest in Phase III which reflected the change in performance during that phase.

The dependant variable was the quality of the proposed strike package (TO3) for each TST, which was determined by expert ratings of the strike package. Specifically, two experts evaluated each strike package. They independently rated the strike package as correct or incorrect with high or low confidence. The ratings were converted as follows: correct with high confidence = 4, correct with low confidence = 3, incorrect with low confidence = 2, incorrect with high confidence = 1, no plan before the TST disappeared = 0, and a plan for striking a target on the no strike list or on the ATO = -1. The expert ratings were the same for 96% of the ratings, and the few disagreements were resolved by conference so that one number was entered into the analysis for each TST.

This measure is comprehensive in that the DTC had to accomplish TO1, TO2, and TO3 in order to get a high score.

The ratings evaluated the plan of the whole team, as opposed to evaluating each individual, making the experiment a single team design. This design is analogous to single person designs for which Anderson (2001) discusses advantages and disadvantages. Anderson notes "In practical affairs, the simple A-B design is sometimes all that is available" (p.313). Single person A-B designs are common in medicine and in behavior modification (e.g. Matyas and Greenwood, 1996). An advantage of these designs is analyzing individuals in depth. Similarly, an advantage of the present single team design is that it enabled the controlled development of a highly experienced team for a task that demanded extended training. A potential disadvantage of A-B designs is that they can confound A and B with the material tested in A and B. We addressed this shortcoming by counterbalancing materials across the previous experiment (Shebilske et al., 2007) and the present experiment. Across the two experiments, therefore, support for the hypotheses cannot be attributed to differences in the materials.

Procedure

Training missions included planning and debrief sessions. During planning sessions, operational teams commit to plans, including the ATO and possible TSTs. During mission execution, the offensive team judiciously adapts plans when unexpected events occur, and the DTC plays the key role of proposing specific adaptations. During debriefs, the team extracts lessons from previous plans and adaptations. The team then incorporates these lessons during subsequent planning, missions, and debriefs. We simulated these operations in the present experiment, each session of which included:

planning (10 min), mission execution (40 min), and debriefing (10 min). To assist them in their work, participants used checklists analogous to those used by the Air Force (Elliott, Cardenas, & Schiflett, 1999).

During Phase I (16 hours), research assistants taught background information using materials from Air Force publications. The background included the characteristics of enemy weapons and friendly assets, communication brevity procedures, and communication dynamics among the team members. The communication dynamics came from an Aptima task analysis. It included why and when teammates communicate with one another and the protocol for who communicates what to whom. Phase I ended with sessions during which the trainees or the instructors could stop the action to address issues. The trainees might stop action to ask questions. The instructors would answer the questions and then resume the action. The instructors might stop action to point out errors. Training before these assisted missions taught what to do. The assisted missions taught how to do it.

During Phase II (11 hours), the trainees performed 11 unassisted sessions. The debriefing included feedback on how well the team was accomplishing its task objectives. Each mission was unique, but they had consistent patterns of differences and similarities, which were created by simulating the same enemy forces attacking with a consistent strategy. The mission difficulty was affected by the number of TSTs per mission and the number of threats that protected the TSTs. Phase II started with warm up trials, which had 3-7 TSTs and 10 threats. Two pre-test missions had 8 and 9 TSTs and 10 threats. These threats were 4 fighters, 1 long rang SAM, and 5 short range SAMS. Each of two post-test missions in Phase II had 8 and 9 TSTs and 37 threats to the strike package. These threats were 24 fighters, 3 long range SAMS and 10 short range SAMS. These enemy threats required the strike package to include enough assets to take out not only the TST but also the threats to the strike package. Difficulty increased gradually over the 11 missions.

During Phase III (9 hr), the trainees performed 9 more unassisted missions, during which they trained for new experiences. The main change at the beginning of Phase III was a strategic variation. For example, during original training, the DTC developed the strategy of using refueling tankers to support their proposed attack packages. This strategy was helpful because the proposed attack packages utilized some assets that were low on fuel due to their execution of operations related to the ATO. A challenging change in enemy strategy might be to have the enemy destroy the refueling tankers, which would force attack assets to return to base to refuel. This requirement would force the DTC to choose different attack package patterns than those to which they had become accustomed. The enemy used its new strategy consistently between the pretest and posttest in Phase III, so that the trainees had an opportunity to practice defending against the new strategy. The number of TSTs (8 and 9) and Threats (37) for the posttest in Phase II was similar to the number of TSTs (9 and 10) and Threats (33 and 35) for pretest missions in Phase III. In contrast, each of the posttest missions for Phase III had 13 TSTs and 45 Threats.

Results

Figure 1shows that TO3 accuracy increased between the pre and posttests in Phase II, fell between Phase II and III, and then increase slightly between the pre and post tests in Phase III. As had been done in the previous experiment by Shebilske et al. (2007), we used SPSS to conducted conservative t-tests that do not assume equal variance. The mean accuracy for TO3 in Phase II increased from 1.5 on the Pretest to 3.2 on the Posttest, as predicted ($t(35) = 3.13$, $p < .01$). Between the posttest in Phase II and the Control pretest in Phase III, the decrease in TO3 accuracy from 3.2 to 1.7 was significant ($t(34) = 2.84$, $p < .01$). In Phase III, TO3 accuracy rose insignificantly from 1.7 on the pretest to 2.2 on the posttest, but the change was not significant ($t(41) = .86$, $p > .05$).

[Insert Figure 1 about here]

Discussion

The pattern of results in the present experiment are consistent with hypotheses 1, 2, and 3, and across the present experiment and the previous experiment (Shebilske et al., 2007) statistical tests support Hypotheses 1, 2, and 3. Statistical tests supported Hypotheses 1 and 2 in the present experiment and Hypotheses 2 and 3 in the previous experiment. The fact that Hypothesis 2 is supported in the both experiments because it is a necessary condition for developing a paradigm to train experienced teams for combat identification in simulated Air Force ISR and DTC teams.

The present experiment also provided pilot data for exploring the training of beginners in this complex domain. Comparisons of our previous and present experiment suggest that the ideal amount of background training using the specific procedures that we developed for teaching beginners what to do and how to do it is closer to 16 hours than 50 hours. Future experiments will systematically investigate the procedures that we are developing for training at both beginner and advanced levels.

In the present experiment, the selection of training scenarios was chosen by the trainer (Figure 2a). The specific sequence, which gradually increased TSTs and then gradually increased threats, was a hierarchical-part-task training protocol (Fredericksen & White, 1989) because a DTC team must detect TSTs before it can design strike packages that take into account threats. Although this sequence is a good fixed one, the present system lacked the instructional intelligence to improve training through adaptability (cf. Bell and Kozlowski, 2002; Freeman, MacMillan, et al., 2006).

Micro-scale analyses and micro-scale interventions, consistently support the hypothesis that adaptive training – training driven by models of instructional strategy – improves learning (e.g. Anderson, Douglass, & Qin, 2004; Anderson, et al., 1995; Conati, C. et al., 1997; Schulze, et al., 2000). We are currently doing research that will expand this predetermined sequence into an adaptive sequence using Partially Observable Markov Decision Process (POMDP) model Figure 2b. Our goal is for this model to provide adaptability with respect to a teams performance in contrast to most formal models in team training which are, ironically, models of individual performers who serve as teammates and/or instructor to a single trainee (c.f., Eliot & Wolf, 1995; Miller, Yin, Volz, Ioerger, & Yen, 2000; Rickel & Johnson, 1999; Freeman, Haimson, Diedrich, & Paley, 2005; Freeman, 2002).

Future experiments will also systematically compare training on the traditional displays that were simulated in the present research with training on new displays that we are developing in other research. The goals of these future experiments will be to facilitate transfer of training from beginner to advanced levels and from traditional displays to new displays.

Future directions will also include replicating this experiment with actual DTC teams. We will collaborate with practitioners to investigate whether simulating new experiences will help actual DTC teams adapt more quickly to real changes. We will investigate hypotheses related to potential learning-to-learn benefits of simulating multiple new enemy forces, new contexts, and new enemy tactics. We will investigate whether learning-to-learn effects for actual DTC teams simulating many new experiences make the DTC teams generally more adaptive to new experiences in operational settings. Ness, Tepe, and Ritzer (2004) reviewed analogous specific and general training benefits to land, air, and naval warfare applications in their book, The Science and Simulation of Human Performance. The present experiment suggests that desktop simulation of new experiences may extend these advantages to combat identification in Air Force ISR/DTC teams.

Conclusion

This research program is distinguished by its logical progression in integrating field research, laboratory research, and field applications. One starting point for the present research was field study of DTC teams in operational settings. Based on these task analyses, we simulated ISR/DTC operations in Aptima's DDD synthetic task environment. The products of this research include a large number of challenging training scenarios that, we believe, should be of direct benefit to Air Force DTC staff training for the field, and in particular training them for a variety of difficult new experiences involving new enemy forces, new contexts, and new enemy tactics.

Accordingly, the present experiment will motivate and guide the next logical steps in this progression, one of which will be to test such training in field settings. This application will have the potential to enhance military simulations with systematic, validated instructional strategies (e.g.Ness, Tepe, and Ritzer, 2004) for ISR/DTC teams especially their adaptation to challenging new missions. Another logical step in both basic and applied future research will be research aimed at understanding effects and interactions among the many aspects of combat identification that occur in this complex domain such as recognizing what an object of interest is, determining whether an object of interest is an expected or unexpected threat or opportunity that demands a response, and identifying friendly assets that could make a contribution to a strike package for a TST without incurring a prohibitive cost of the asset not being available to execute the ATO.

References

Anderson, J. R., Corbett, A. T., Koedinger, K. R., and Pelletier, R. (1995). Cognitive tutors: lessons learned. *The Journal of the Learning Sciences,* 4(2), 167-207.

Anderson, J. R., Douglass, S. & Qin, Y. (2004). How should a theory of learning and cognition inform instruction?. In A. Healy (Ed.) *Experimental cognitive psychology and it's applications*. American Psychological Association; Washinton, D. C.

Anderson, N.H. (2001). *Empirical Direction in Design and Analysis.* Mahwah, NJ.: Lawrence Erlbaum Associates.

Bell, B. S., & Kozlowski, S. W. J. (2002). Adaptive guidance: Enhancing self-regulation, knowledge and performance in technology-based training. Personnel Psychology, 55, 267-306.

Conati, C., Gertner, A., VanLehn, K., and Druzdzel M. (1997). On-line student modeling for coached problem solving using Bayesian networks. In A. Jameson, C. Paris and C. Tasso (Eds.), *User Modeling: Proceedings of the Sixth International Conference, UM97,* 231-242. Vienna: Springer.

Elliott, L. R., Cardenas, R., & Schiflett, S. G. (1999). *Measurement of AWACS team performance in distributed mission scenarios.* Retrieved August 28, 2003, from http://www.dodccrp.org/1999CCRTS/pdf_files/track_3/013ellio.pdf

Eliot, C., & Woolf, B.P. (1995). An adaptive student centered curriculum for an intelligent training system. *User Modeling and User-Adapted Instruction*, 5, 67-86.

Freeman, Jared. (2002). I've got synthers. Who could ask for anything more? *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society.* Baltimore, MD.

Freeman, J., Haimson, C., Diedrich., Paley, M., (2005). Training teamwork with synthetic teams. Clint Bowers, Eduardo Salas, & Florian Jentsch (eds.), *Creating High-Tech Teams: Practical Guidance on Work Performance and Technology.* Washington, DC: APA Press.

Freeman, J., MacMillan, J., Haimson, C., Weil, S., Stacy, W., and Diedrich, F. (2006). From gaming to training. Society for Advanced Learning Technology. Orlando, FL. 8-10 February 2006. 8-10February 2006.

Matyas, T. A., & Greenwood, K. M. (1996). Serial dependency in single-case times series. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and* *analysis of single-case research* (pp. 215–243). Mahwah, NJ: Lawrence Erlbaum Associates.

Miller, M., Yin, J., Volz, R.A., Ioerger, T.R., Yen, J. (2000). Training teams with collaborative agents. *Proceedings of the Fifth International Conference on Intelligent Tutoring Systems*, (ITS-2000), 63-72.

Ness, J.W., Tepe, V., and Ritzer, D.R., (2004). *The Science and Simulation of Human Performance*. Amsterdam: Elsevier.

Schulze, K. G., Shelby, R. N., Treacy, D. J., and Wintersgill, M. C. (2000a). Andes: A coached learning environment for classical Newtonian physics. In J. A. Chambers (Ed.) *Selected Papers from the 11th International Conference on College Teaching and Learning,* 151-162, FCCJ, Jacksonville, FL.

Shebilske, W., Gildea, K., Freeman, J., & Levchuk, G. (2007) Training Experienced Teams for New Experiences. Paper presented at the 51st annual meeting of the Human Factors, Society, Baltimore, MA.

Figures

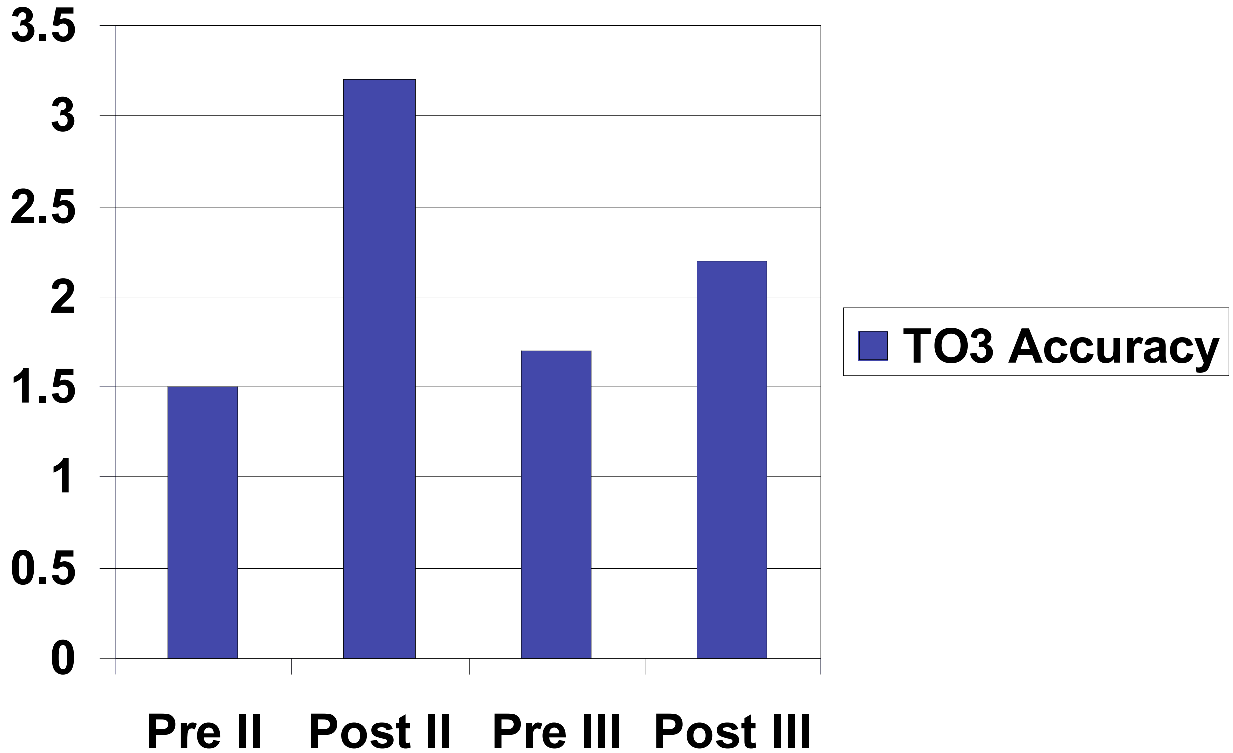

Figure 1. Mean accuracy ratings for training objective 3 (TO3) for two sets of pretests and posttests: Phase II training (II) and Phase III Training (III).

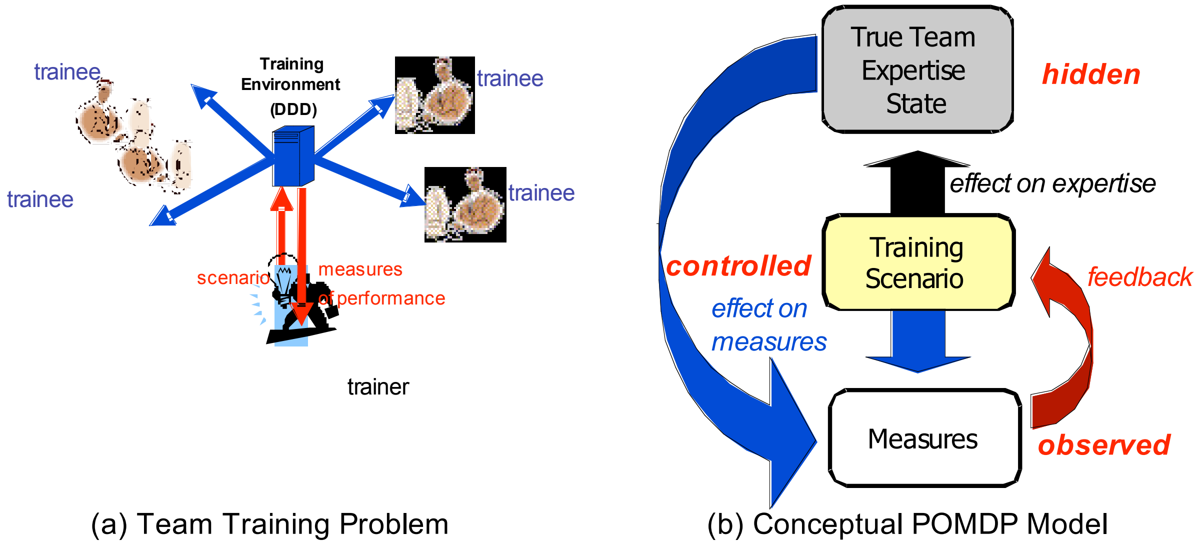

Figure 2: The Problem of Training and Conceptual POMDP Solution