Training for Crisis Decision-Making: Psychological Issues and Computer-Based Solutions

JANET A. SNIEZEK, DAVID C. WILKINS, PATRICK L. WADLINGTON, AND MICHAEL R. BAUMANN

JANET A. SNIEZEK is a Professor of Psychology, Business Administration, and the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign, has held faculty positions at the University of Chicago, Cornell University, and Purdue University, and has been a Visiting Scholar at Stanford University. Dr. Sniezek's research interests include advice giving and taking, behavioral decision-making with collaborative technologies, expert system evaluation, and performance assessment. Her research has been published in numerous journals, including *Organizational Behavior and Human Decision Processes, Journal of Behavioral Decision Making,* and *International Journal of Forecasting.*

DAVID C. WILKINS is an Associate Professor in the Aviation Institute and Beckman Institute, Director of the Knowledge Based Systems Group, and departmental affiliate in Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign. He received his Ph.D. from the University of Michigan in 1987. His research interests are in knowledge-based systems, machine learning, human-computer interaction, and intelligent tutoring.

PATRICK L. WADLINGTON is a graduate student in industrial/organizational psychology at the and an affiliate of the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign. His research interests include psychometrics, situation assessment, performance measurement, selection, and training.

MICHAEL R. BAUMANN is currently an Assistant Professor of Psychology at the University of Texas at San Antonio, and previously held a faculty position at Washington State University. He received his Ph.D. in psychology from the University of Illinois at Urbana-Champaign in 2001. Dr. Baumann's research interests include decision-making under acute stress, group performance, and the interaction between cognition and emotion.

ABSTRACT: Crises demand swift and effective decision-making; yet there are many problems in training personnel on the skills necessary to achieve the goals of crisis management. This paper has three objectives concerning training for crisis management. First we integrate diverse literatures and present a framework for an understanding of the unique challenges in crisis management training, and the role of training systems with capabilities for simulation, immersion, and critiquing. Second, we describe an example of a trainer for ship damage control, called DC-Train, which ad-

Journal of Management Information Systems / Spring 2002, Vol. 18, No. 4, pp. 147–168.
© 2002 M.E. Sharpe, Inc.
0742–1222 / 2002 $9.50 + 0.00.

dresses these challenges. This system consists of a first-principles simulator that generates large numbers of realistic scenarios, an immersive multimedia interface that helps elicit psychological processes involved in actual crisis management, and a critiquing expert system that provides real-time and post-session feedback on human decision-making performance. Finally, we present an empirical method for evaluating the effectiveness of such a system for crisis management training. Results of evaluation experiments with participants in a ship damage control training program indicate that the described computer-based trainer has psychological realism and improves decision-making performance.

KEY WORDS AND PHRASES: artificial intelligence, computer-based training, crisis management, human-computer interaction, human resource management, ship damage control.

> One very specific environment in which decision making is critical is a crisis: An unexpected, life-threatening, time-compressed event, such as an engine failure or a frozen flight control that usually occurs once in a lifetime. [37]

THE ABILITY TO MAKE GOOD DECISIONS in the midst of a crisis is both extraordinarily important and extraordinarily difficult. Analyses of disasters often point to mistakes that humans make that worsen, fail to prevent, or cause disasters. For example, in the aviation community, an industry-wide analysis has shown that over 70 percent of aviation accidents can be attributed to human error [39], and in medicine it is claimed that between 70 percent and 82 percent of anesthetic incidents in the operating room have been attributed to human error [12]. Errors in human actions can have widespread disastrous consequences, such as the bombing of an embassy, the shooting down of a civilian airliner, or meltdown in a nuclear power plant.

In some domains, conventional noncomputer-based training approaches have serious limitations with respect to training personnel for crisis management duties. Many of these limitations can be reduced with the use of advanced training technologies that include three components: an immersive multimedia interface, a crisis simulator, and a critiquing system. In this paper we identify some of the theoretical and practical issues in developing training programs for crisis management, and provide an illustration of a training system for crisis management in the domain of ship damage control. In the first section, we draw on diverse literatures to create an integrated framework for understanding the special problems in training for crisis management. This framework is then used to show the advantages of training with a system with an immersive interface, crisis simulator, and critiquing system compared to only "conventional" training (that is, noncomputerized systems with none of these features). The remainder of the paper is devoted to the case of training for ship damage control. We discuss the special benefits of DC-TRAIN, a multimedia immersive trainer with crisis simulation and a critiquing system, and provide empirical data on its training effectiveness for crisis decision-making.

Training Issues and Solutions

Crisis Management

CRISES COME IN MANY FORMS: a devastating earthquake, hostile takeover of a company, airplane crash, food or fuel shortage, taking of hostages, stock market crash, or enemy attack. Regardless of the domain, crises usually share four features: uncertainty, rapid onset [15], imminent or realized severe losses, and a lack of controllability.

There are many sources of uncertainty for those who attempt to influence the events of a crisis; the cause of the problem and its extent or duration may be unknown or unknowable at the time of the event. More importantly, it is difficult to know how to proceed to manage the crisis. To take effective actions, it is necessary to know the likely course of multiple events and how to allocate resources in dealing with them. Although some of the uncertainty results from the rapid onset of the crisis, some of it is due to the inherent complexity [54] and unpredictability of the events comprising the crisis.

Of course, a crisis may be long in the making, but the point at which there is consensus that a crisis exists is typically characterized by a rapid unfolding of events that require swift action in response. The goal of the response is to prevent or reduce negative consequences. Surprising events that are difficult to influence are labeled crises only if it is apparent that something of value is being lost, or soon will be.

Despite the urgency, effective action is difficult. Typically the losses are irreversible and, to some extent, unpreventable. Although the events of the crisis cannot be completely controlled, they can be partially influenced. Crisis management is the set of actions taken to exert control over the events of a crisis to minimize losses. It is similar to risk management—except that the events are real, not potential, and action takes precedence over planning. To be effective, crisis management involves making good decisions [40] under severe time pressure and uncertainty. However, the very factors that defines crises make decision-making all the more difficult. For example, time pressure reduces the quality of human judgment [6, 21]. In a recent study of crisis management by Air Force commanders [1] suggest that more experienced commanders were more resistant to the negative impact of time pressure and better at processing information and consistently following a plan. One means of increasing the experience level of personnel is through training.

Challenges in Crisis Management Training

Crisis management is a highly complex skill that is difficult to acquire for many reasons. First, crises are, by definition, rare, making it difficult to acquire direct experience in the management of crises in a given domain. Further, experience is not always a good teacher [10]. The highly uncertain and complex environment of a crisis can be a poor place to try to discern cause-effect relations. Second, when crises do occur, conditions are unfavorable for training [17]. Those with relatively more expertise must act and have no time to attend to the needs of a trainee. It may not even be possible for novices to be in a position to passively observe the decision-making of experts during the crisis. Third, crisis management skills may not generalize well across crises. The complexity and uncertainty of the environment in which crises occur mean that each crisis is unique. The scenario of a disaster is rarely the one for which personnel were trained. In addition, the skills and knowledge required need to be updated continuously both because the environment of crises is dynamic, and because the regulations and technology for managing crises continually change. In short, the major problem you face in trying to learn crisis decision-making through experience is that it is rare to experience a crisis, and even if you do, you may not be able to learn from it. Then there is the problem that even if you do learn, you may learn some wrong things, or things that are useless in the next crisis.

Thus it is not surprising that structured training programs are often used to foster the acquisition of crisis decision-making skills. These exist in many forms, and several diverse literatures address topics relevant to crisis management training. We proceed toward the goal of synthesis by examining theoretical and practical issues involved in developing a training program for crisis decision-making, taking the perspectives of human resource management and psychology.

A Framework for Training Issues

A major problem at the start of establishing a crisis management training program is conducting a needs assessment to identify what needs to be taught [25]. Domain experts are an important resource in this endeavor; however, reliance on them entails problems in recruitment and selection. Those with experience with past crises are an important source of information, although the aforementioned problems limit the number of such experts who exist and the currency of their expertise. Further, any one person's direct experience with crises will be limited. Even if multiple persons have input on the content of the training, there is still the problem of disagreement among them, and the fact that none of them can predict the exact nature of a future crisis. This may lessen respect for the training, promoting an attitude that "they cannot teach you what you need to know." Although no one can know exactly what needs to be taught, it is likely that both domain knowledge and procedural knowledge as well as general problem-solving strategies are crucial to crisis management. The latter requires a wide range of skills (such as, communication, social, perceptual, and decision-making) that further complicates the process by widening the scope of training.

Work in many organizations is increasingly performed by groups or teams of people. Crisis management is particularly reliant on interactions among members of teams, posing the problem of how to best conduct team training. Ideally, all members who will perform as a team would be trained together. This can be difficult when each member needs individualized, highly specialized training for their own responsibilities. It also may not be practical to train everyone together if some persons have significant high-level responsibilities outside interacting with those in the group.

Once the content of a training program is determined, there is the problem of assessing how well the trainee has mastered it. There are two specific administrative goals. One is measuring performance for purposes of personnel decisions, such as, deciding if the trainee has successfully completed the training, needs remedial training, or is the best of the training group [23]. Second, performance assessment data provides important input for utility analyses to determine the value of a training program [40]. Although they are more difficult to develop and use, measures of process are more useful in diagnosis of individual performance or the training program as a whole.

Another purpose of obtaining information about performance is to give the trainee feedback that can be used to improve performance during training. To be most effective, feedback needs to be prompt and specific, and go beyond communicating outcomes or error. The information given to the trainee should provide sufficient explanation for the trainee to develop an improved understanding of events in the crisis, and their connection to actions taken or not taken. Generic instructions that repeat rules are generally not as beneficial as instructions directed at the particular deficiency displayed by the trainee. In addition to improving understanding, explanations allow users to verify or supplement their own explanations and reconcile contradictions between their expectations and system recommendations [36].

Delivery of feedback requires some form of interaction between the trainer and trainee. In its simplest form, communication is unidirectional, from the trainer to the trainee, and occurs only after execution of actions in a crisis scenario. Cost considerations often confine interactions to a single trainer communicating with multiple trainees.

The realism of training is one of many issues pertaining to the transfer of training. But given its special significance for crisis management training, we are treating it as a separate topic. Some of the forms of training that are least expensive, such as classroom training and written materials, have the least realism. The most realistic training takes place in contexts that represent the context of an actual crisis. If done correctly, realistic training is more beneficial than classroom training due to the advantages of active training over passive training [3]. A lack of realism can prevent transfer of training (such as, for performance under time pressure [56]). The exact requirements for realistic training vary across domains but usually involve some combination of natural and simulated features [1]. For financial and ethical reasons, it is not possible to involve real events that lead to real losses. Thus, training contexts often make extensive use of simulations that approximate crisis events. Yet, even very expensive simulations that look like crises may not be sufficiently realistic [14, 41].

For training purposes, a simulation is realistic if it induces the same *psychological processes* in the training context that are experienced during an actual crisis [14, 32]. Perhaps the greatest challenge in training for crisis management is inducing the psychological processes associated with the acute stress experienced in actual crises. Acute stress is a state that occurs in situations of potential harm, time pressure, and arousal [42]—all characteristic of crisis management. Although there has been a longstanding interest in the problem of preventing decrements in human performance under stress, research on human performance under acute stress is limited. It is difficult to conduct research on human performance in crises for the same aforementioned reasons it is difficult to train. However, there is evidence that cognitive impairment during acute stress may be more serious for some personality types than others [24]. Recent theory proposes that the manner in which individuals evaluate the level of potential harm, time pressure, and demands of the task plays a key role in how well the individuals perform [4]. According to Chi et al. [11], the "training and simulation" intervention strategy has the goal of changing the appraisal process in the face of a crisis. The specific recommendation is to "provide the individual with the knowledge and confidence that will lead to the development of positive performance expectations and result in more effective task performance under stress" [11, p. 187]. It follows that one way to promote effective crisis management under acute stress is to train individuals to perform under acute stress, or at a minimum, some combination of arousal, time pressure, and anxiety.

A final problem that needs to be addressed with any training program is transfer of training. The motivation for realistic training is to allow for the skills acquired in training to be applied in the context of an actual crisis. There are other factors that can prevent transfer of training for crisis management. One is that the nature of crises can change over time due to changes in the physical, social, or economic environment. In addition, crisis management technologies and procedures change. In short, one problem is that today's training may not help tomorrow's performance. A goal that is not easy to realize in training for crisis decision-making is the provision of a sufficient number of novel crisis scenarios to prepare trainees for crises that never happened, but could.

Training with Immersive Interfaces, Simulators, and Critiquing Systems

An alternative or supplement to conventional training methods is the use of advanced computer-based training technologies. This approach involves use of a training program that is initially based on established specific doctrine used by an organization [34]. Computer-based systems to facilitate human learning exist in many forms, but here we are interested in those that combine three features: a simulator for generation of training examples, an immersive interface to assist with replicating the information overload associated with many crisis situations, and a critiquing system for generation of the feedback that is normally given by a teacher. In Table 1, we present a framework for understanding the major issues in training for crisis management, and the ways in which an expert critiquing and simulation system can address the problems associated with conventional training approaches.

An obvious advantage of an expert critiquing system is that it can alleviate the problem of a shortage of human experts in training. Subject matter experts are still essential; they provide the necessary knowledge for development of the critiquing system and crisis simulator. The human experts are also valuable in testing and vali-

Table 1. A Framework for Understanding Training Issues in Crisis Management and Solutions Based on an Expert Critiquing and Simulation System

Training issue	Problem for crisis management training	Immersive interface, expert critiquing and simulation system solution
Expert selection and recruitment	Few experts; available experts have limited experience with actual crises; lack of agreement among experts; instructor-centric training	Critiquing system makes use of expertise of multiple experts; standardization of training; trainee-centered training
Training content	Difficult to define precise domain and procedural knowledge; wide variety of skills is needed	More comprehensive knowledge base; simulation capability allows training multiple skills
Team training	Difficult to train in teams of specialists	Team members represented by intelligent agents
Effectiveness assessment	Extra effort to specify and measure performance, use of subjective measures; emphasis on outcomes over process	Effective performance can be well-defined and measured by critiquing system; emphasis on reasoning process
Feedback	Multiple acceptable courses of action; need to tailor feedback to trainee's mental model; time delays	Can assess trainee actions automatically; generate individualized critiques
Interactions with trainer	Cost encourages many trainees per trainer and limits two-way interaction	Efficient two-way one-on-one interaction during performance; advising to encourage development of mental model
Time for training	Repeated periods of substantial time needed for learning of complex skills	Flexible scheduling for massive practice; training anytime and anywhere computer access is available
Cost	Costs increase with number of trainers, length of training, and as training context approaches that of actual crises	High initial development costs; predictable implementation and maintenance costs
Realism	Actual crises impossible for training; difficult to create acute stress	Simulation allows for representation of actual crises; immersive interface contributes to acute stress
Transfer of training	Expert knowledge and experience may not generalize to future crises; difficult to prepare for uncertainty of crises	Multiple novel scenarios increase breadth of knowledge, aid in the reduction of uncertainty

08 sniezek.p65 153 01/19/2002, 10:47 PM

dating the system. The real contribution of the expert critiquing and simulation system to the problem of a shortage of human experts is realized once the system has been perfected. Multiple versions of the computer-based trainer allow many students to be trained simultaneously. Further, this replication provides standardization that is usually not present with training from multiple experts in a complex domain. As is true with the adoption of many forms of technological trainers, a shift in emphasis occurs in who and what is at the center of the endeavor. Although conventional training is instructor-centric and emphasizes teaching, training with an expert critiquing and simulation system is trainee-centric and emphasizes learning [25].

Whether training in crisis management is conducted by the experts themselves or with a computer-based system, it is subject matter experts who provide the domain and procedural knowledge that will be the context of training. But in building a computer-based system, expert knowledge is elicited from multiple subject matter experts, thereby combining their respective inputs into what can be a more comprehensive knowledge set than any individual expert possesses. This helps designers create a system that can simulate a large number of crisis scenarios that require the trainee to engage in actions requiring a wide variety of skills, such as recognizing patterns, detecting signals, coordinating with team members, communicating effectively, predicting outcomes of events, making decisions, and remembering protocol. The comprehensive knowledge base and ability to simulate large numbers of scenarios requiring complex skills greatly expands the scope of training for crisis decision-making.

A simulator provides the opportunity to represent many aspects of decision-making in crises. This potential has not been realized in terms of representing other members of a team with whom one must interact in a crisis. Advances in the ability to simulate team behaviors [13] allow sophisticated means of conducting "team" training. Intelligent agents can represent personnel, and advanced models can allow a wide range of human behaviors—good and bad—to be represented in crisis scenarios.

An expert critiquing system has significant value in that it can analyze a simulated crisis and the actions taken to control it. This means that it can also assess the trainee's performance according to well-defined criteria. Although a performance analysis can be done with conventional training techniques also, it requires an extra effort that is rarely undertaken with the same care as is put into the design of the expert reasoning system. Conventional training makes typical use of subjective methods of performance assessment, for example, by having the trainers rate trainee performance. In other cases, it tends to be focus on outcomes—which are easily observed—rather than process. In crisis management, outcomes are a particularly misleading indicator of performance. Losses are almost always guaranteed, but do not necessarily reflect poor crisis management. Performance assessment with an expert critiquing system concentrates on the quality of the reasoning process used in managing the crisis [43].

Although an expert critiquing system can yield optimal decision strategies, this does not by itself guarantee that students will learn from it [13]. The real benefit of a critiquing system can be seen in performance feedback supplied to the trainee, a key feature of many intelligent tutoring systems [45]. It is important that trainees understand their deficiencies in training exercises [16]. Human experts can tailor feedback to individual trainees, but it is a very labor-intensive process. The expert must understand the reasoning of the trainee as well as the events of the crisis and the actions taken by the trainee in response. An expert system can do this analysis automatically, in real time, allowing for swift feedback. Speed, of course, is not enough. One important feature of this feedback is that it emphasizes the trainee's process, rather than the outcomes only. It is well-known in human decision-making research that process feedback is superior to outcome feedback [3]. A potential problem with feedback is that if it is counterintuitive, trainees may reject it and not improve in performance. System users seek explanations to resolve unexpected results. By including explanations in critiques, feedback is more likely to be understood and accepted. Explanations also increase trust, which in turn promotes acceptance of feedback [36, 47, 55].

A model of the effects of feedback suggests that different types of feedback are required at different levels of expertise [33]. Feedback is assumed to focus attention on levels in a hierarchy of goals. Feedback that focuses attention on lower levels of the hierarchy is best for initial learning on complex tasks, but as the individual's skill increases, feedback focusing on overall task performance will be more useful. An expert critiquing system can provide feedback about performance at any level at the time it is most valuable, and match the nature of the explanation to the experience level of the user. This is especially advantageous given that persons with more experience appear to have systematically different patterns for wanting explanations than do those with less experience [36].

The instructional process is further enhanced if there are various means for communication of feedback and other performance information. An expert critiquing system is very flexible in terms of the timing and nature of interactions with the trainee. An expert critiquing system can analyze events and actions as they occur; thus, there is no need for any delay in providing information to the trainee. Further, communication to instruct and improve trainee performance is bidirectional and individualized. Two-way interactions that allow either party to initiate communication at any time have particular advantages; the expert can choose to deliver advice or feedback, or the novice can request such assistance. There are occasions in which it is better for the trainee to receive advice and then practice applying it than to practice with faulty actions and subsequently try to correct them. Human tutors vary in their techniques and not all of them consistently use the most appropriate teaching techniques. In contrast, an expert system can ensure that specific beneficial methods (such as, eliciting self-explanation [11], asking questions [26], or giving hints and suggestions [31]), are used when they are determined to be most appropriate.

Another collection of problems has to do with the simple questions of the time and place to do the training plus choice of what materials are needed for good training. If written correctly, the expert critiquing and simulation system can be a self-contained training system. This characteristic allows training to occur at any place and any time. The only material needed is a moderately high-functioning computer. In essence, especially in today's day and age, this makes the training system flexible [38]. For example, if so desired, an employee could train in the comfort of home. Any feature that increases flexibility removing barriers to practice—all else being equal—will increase actual training time. Opportunities for practice allow the trainee to realize the benefits of critiquing. Understanding one's deficiencies with a skill and then being able to practice the skill in context is necessary for effective training [16].

As for the new problems that arise with the high-tech approach, cost is definitely the largest. The initial research and development costs of an expert training system with extensive simulation capabilities is prohibitive in many domains. It can be unaffordable and unwise to modify a system from another domain. If the initial costs of system development can be met, there is the advantage that implementation and maintenance can be predicted with reasonable accuracy. Finally, although cost is a definite obstacle in the present, it may prove to be trivial in the near future. Cost is likely to be reduced with the further advancement of technology, both in terms of hardware and software and in development costs for expert training systems.

A final, but extremely valuable benefit of a powerful simulator that presents scenarios to the trainee through an immersive interface is that it greatly improves the chances for transfer of training. The simulator creates an active learning environment that is much preferred to passive ones [34, 35, 51], which is advantageous in any training. But, there are several reasons that such a simulator is particularly valuable in training for crisis decision-making. First, the simulator can elicit the psychological processes of crisis management in actual crises, making training more realistic. In particular, simulations can create acute stress through multimedia interfaces, time pressure, information overload, and graphic representation of the most disturbing losses of a crisis. Although support systems can be effective in improving task conditions, for example, by reducing information load [27] the burden of processing information will remain high in a crisis. This means that representation of crisis conditions may require design features to increase information load. Another valuable feature of a powerful simulator is that it can create a large number of novel scenarios. As a trainee's exposure to total number of scenarios and number of different types of scenarios increases, the trainee's ability to cope with uncertainty and transfer training to a real crisis improves. Experience managing many kinds of crisis scenarios reduces the chance that an actual crisis will be so novel as to render all training experience irrelevant.

The Case of Ship Damage Control Training

CRISES ON SHIPS, BY THEIR VERY NATURE, pose the problems addressed in Table 1 in the training of personnel responsible for making decisions during the crises. In this section we report on DC-TRAIN, a computer-based training system that addresses these problems in the context of training for ship crisis management. DC-TRAIN is designed to train and measure the performance of Damage Control Assistants (DCAs), the naval officers responsible for managing crises on their respective ships. After describing specific features and capabilities of the system, we report data from empirical evaluations of it with DCA trainees at the U.S. Navy's Surface Warfare Officer School (SWOS) in Newport, Rhode Island.

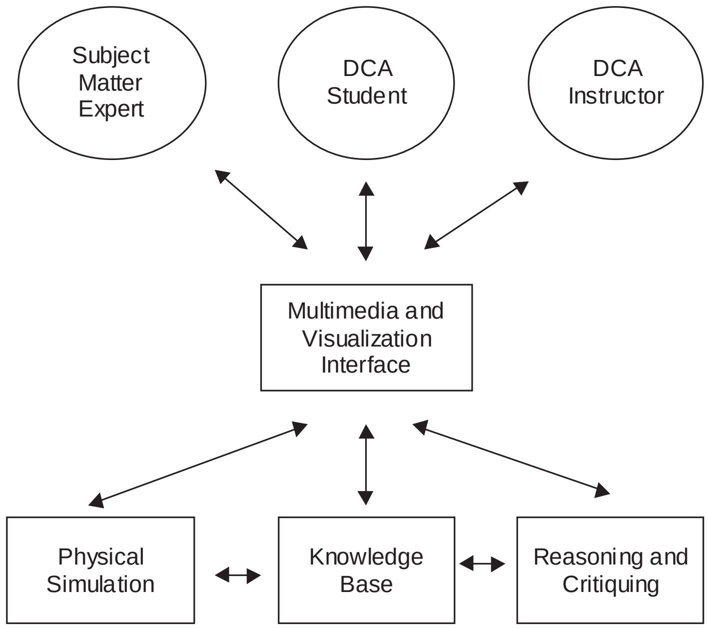

Figure 1. DC-TRAIN Major Components

The DC-TRAIN System

DC-TRAIN is an immersive whole-task simulator-trainer under development at the University of Illinois and under validation at SWOS [9]. During the course of a simulated crisis scenario, the DCA makes decisions directing the repair and containment of damage during crisis situations. DC-TRAIN provides the facilities to specify and simulate from physical principles a large number of damage control situations. The spread of damage, the effectiveness of damage control actions, and ship personnel other than the DCA are simulated. Figure 1 illustrates the relationship between the major components of DC-Train and the users of the system.

Two of the immersive interface components of DC-TRAIN 2.5 are illustrated in Figures 2 and 3. In Figure 2 the window marked A is the command list, where the user selects actions to perform. Window A is modeled after the screen layout of IDCTT 3.0, a system developed by Tekamah and NAVSEA. In Figure 3, Windows E and F feature the graphic visualization module displaying two different views of the ship, and during the run of a simulation the user may open as many different views as desired. In window D, audio and video feedback from the system is presented. The message blank window in area C is used to gather command parameters from the user. Window D presents system critiques of the user's actions.

A scenario generator module in DC-Train allows an instructor and students to specify crisis scenarios [44]. The primary information that needs to be specified is the initial ship crisis state, including the location, size, and initial time of crisis events like fire, smoke, flooding, pipe rupture, hull rupture, or wall rupture. The initial state of the

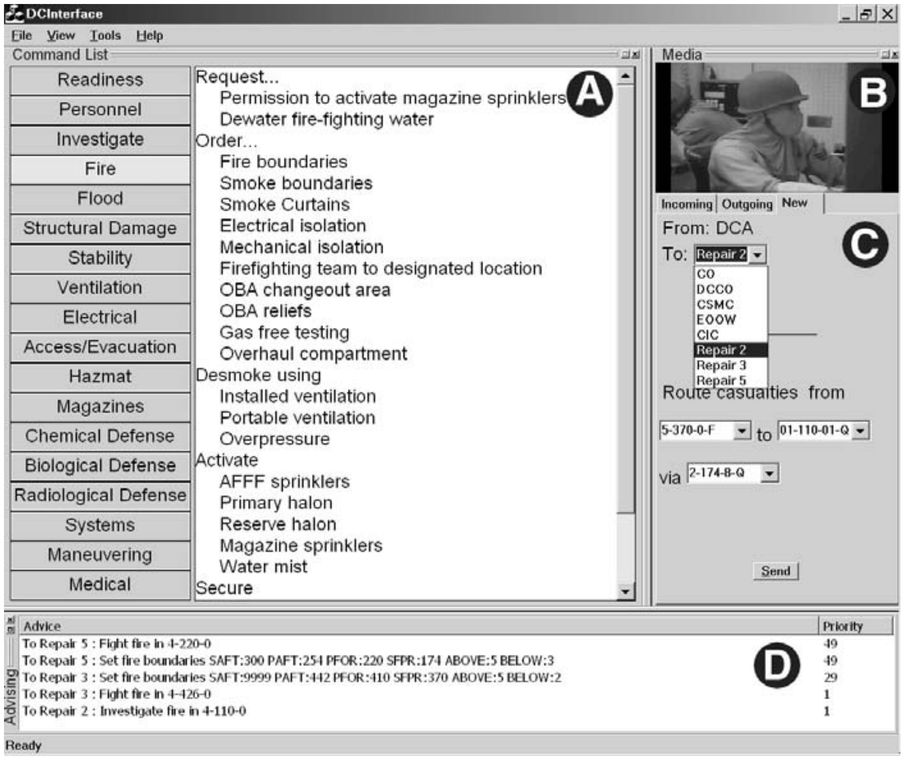

Figure 2. DC-TRAIN Graphical User Interface (GUI) During a Crisis Scenario

ship can also be specified such as the state of doors, hatches, fire pumps, the amount of material in compartments, and the like. Consequently, an enormous number of novel scenarios with widely varying features can be simulated.

A physical ship simulator module in DC-Train propagates the consequences of the initial damage through time [44]. The simulator contains numerical algorithms for simulating the spread of the damage control events that are specified by the scenario specification interface. For example, it models the spread of fire and smoke, flooding, fire main rupture, hull rupture, the pressure at any point in the fire main, and the time to engulfment of a compartment by fire. The physical ship simulator models the suppression of the crisis events by automatic systems such as halon, and by the actions of human personnel such as by the use of fire hoses. Interactions between crisis events are modeled, such as the effect of flooding a compartment on a fire that is in the compartment. The simulator capabilities mean that many physical events of ship crises are well-represented in any given scenario.

An intelligent agents module in DC-Train simulates the various personnel on the ship involved with a damage control crisis. The intelligent agents model humans in charge of various tasks, allowing training on team skills. For the most part, the intelligent agents correspond to entities that are in direct communication with DCA, who is the decision-maker being trained by DC-Train. Ship personnel that are simulated

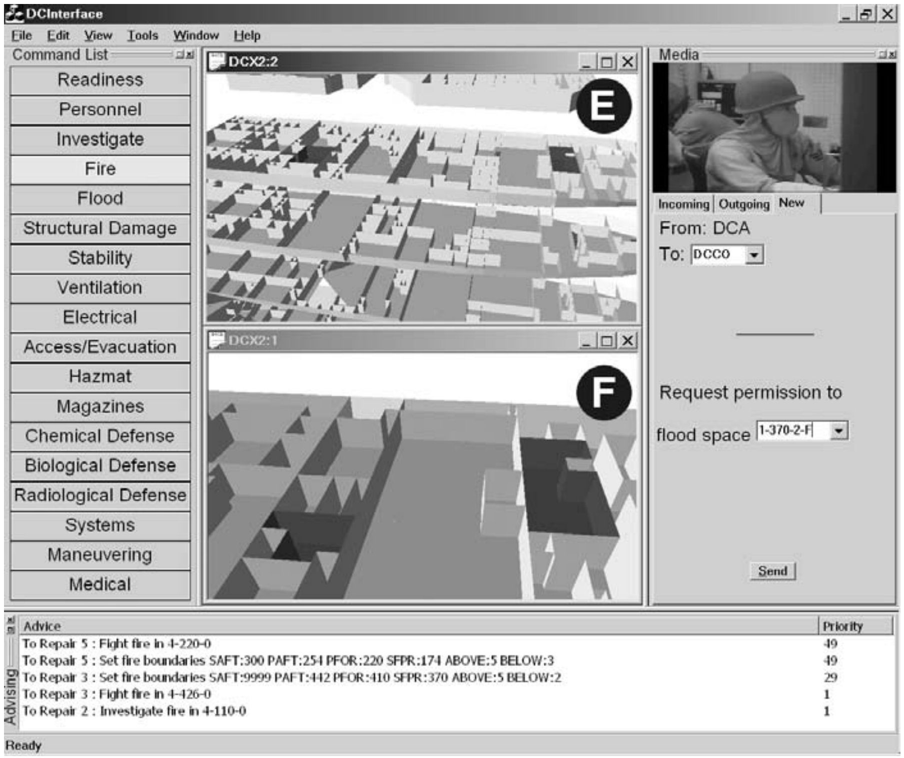

Figure 3. Another View of the DC-TRAIN Unified GUI, Which Highlights the 3-D Interactive Graphical Visualization in Windows E and F

include the Bridge/Commanding Officer (Bridge/CO), the Engineering Officer on Watch (EOOW), the Damage Control Console Operator (DCCO), the Combat Systems Maintenance Central (CSMC), and the Repair Locker Leaders (RLL), and the Repair Locker Investigators. The DCA can send communications to the agents via the command interface. The responses of the agents are received via the audio/video interface and text-based interface.

An automated DCA module in DC-Train allows the student to run DC-Train and observe the actions of the automated DCA in solving a crisis scenario [8], providing additional flexibility in training techniques. In empirical evaluations involving hundreds of scenarios, the automated DCA has been found to solve crises at the expert level. Two different versions of an automated DCA have been implemented. One version of the automated DCA takes a traditional AI approach that contains knowledge in the form of plans, rules, preconditions, and constraints [7]. This approach probabilistically ranks alternative actions and thus can evaluate a DCA's behavior when actions are taken that are not the same as the top action that the automated DCA concluded was the best one to take at a particular point in time. The other version of the automated DCA is a Petri network approach that ranks the value of alternative actions by hyper-simulating each of their consequence on the state of the ship using a time horizon of 10 or 20 minutes [8]. The Petri network approach is fast enough to simulate the alternative actions into the future in real time.

The automated DCA module is also used as the foundation for a critiquing module [8]. It compares students' actions to the actions suggested by the traditional AI version of the automated DCA, which is used as a gold standard for performance. By this comparison, it can provide feedback in terms of errors of commission (actions the DCA took but should not have taken) and errors of omission (actions that the DCA did not take but should have). The student's performance is being evaluated in terms of whether or not the student made the best decision given the information available. The comparison is between an "expert" decision and the student's decision, and is not influenced by probabilistic elements regarding firespread, and so on, or the uncertainty regarding the actual state of the ship at any given time. Thus, the critiquing module can thereby yield a process measure of trainee performance.

The critiquing system can also use a Petri network envisionment to judge the correctness of a "novel" student action (that is, one not in the automated DCA's knowledge base). If the Petri network simulations reveal that the novel action would have a beneficial effect on the course of the crisis when the crisis is simulated 10 or 20 minutes into the future, it would be judged to be a good action. How good can be determined by comparing the simulated outcome to the outcome for the best action simulated by the Petri network. The advantage of this approach is that the accuracy of the critiquing is not affected by the amount of knowledge in the automated DCA. In either case, the expert system and DCA have access to the same information regarding the state of the ship. The traditional AI approach computes the best action given the information available to the DCA. Likewise, the Petri network approach simulates whether the action would be beneficial given the information available to the DCA. Thus, the decision can be evaluated in terms of whether it was a good decision given the information that was available to the DCA at the time instead of in terms of the decision's eventual effect on the state of the ship. Therefore, both feedback regarding decision quality and critiquing are linked to the quality of the decision and not the final outcome.

Of course, the fact that a training system is based on sound principles does not necessarily make it effective. A crucial part of the DC-TRAIN research project is the testing of the system with DCA trainees.

Empirical Evaluation of DC-TRAIN

Assurance that a computer-based training system is well built in the sense that is reliable, fast, and so forth, is necessary but not sufficient to conclude that it is an effective trainer. The real proof is in the performance of the trainees. With this assumption, we conducted a series of empirical tests of DC-TRAIN with DCAs at SWOS. Although there are many criteria of interest in evaluating the training effectiveness of any system, we concentrate here on two measures that relate to particularly important but difficult problems with regard to training for crisis decision-making: human per-

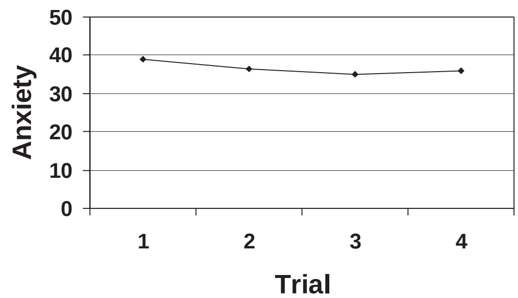

Figure 4. State Anxiety Scores Over Trials

formance and system realism. We include some discussion of measures relevant to other criteria for training effectiveness, such as, user satisfaction.

The first study to evaluate the training effectiveness of DC-TRAIN version 1.0 involved the assessment of performance and other measures over training trials. The study also included the experimental manipulation of the number of training trials prior to performance on a criterion scenario. The criterion consisted of completion of a trial on IDCTT 1.0, a computer-based system with a high fidelity, but the ability to generate only one scenario that focuses on fire spread. A total of 58 DCA trainees completed three training trials with DC-TRAIN as well as a single trial with IDCTT. To experimentally manipulate training experience on DC-TRAIN, participants were scheduled in such a way that approximately one-fourth of them did IDCTT first, one-fourth did IDCTT after one trial of DC-Train, one-fourth after two trials of DC-Train, and one-fourth after three trials. The design was a 4 (DC-Train timing) by 4 (trial) with the first factor between-subjects and the second within-subjects. The participants completed various measures (described below) after each trial in the role of DCA as well as at the end of the simulator data collection. Before discussing the human performance data, we describe the data on these measures.

On all four trials, subjective difficulty, subjective time pressure, and expended effort were assessed via ratings. Subjective difficulty was measured by asking participants how difficult the task was on a scale from 1 (not at all) to 9 (extremely). Subjective time pressure and effort data came from the relevant subscales of the TLX (Task Load Index) measure of subjective workload, in which 11-point scales ranged from "low" to "high" [29]. Mean ratings for each of these measures were significantly above the midpoint of their scales ($p < 0.02$) and did not decline over trials ($p > 0.48$).

Theory on the role of anxiety in performance in crises indicates that the presence of substantial anxiety is a desirable feature of a simulator for crisis management training [5]. State anxiety as well as trait anxiety was measured with the STAI [48]. Compared to the norms for working adults aged 19 to 39, this sample had state and trait anxiety scores equivalent to those at the 80th and 43rd percentiles, respectively [49]. State anxiety, illustrated in Figure 4, was on average 20 percent higher than the baseline given by trait anxiety scores and significantly higher than predicted by trait anxiety on trials 1, 2, and 4 (all $p < 0.02$) and marginally greater on trial 3 ($p < 0.06$). Further-

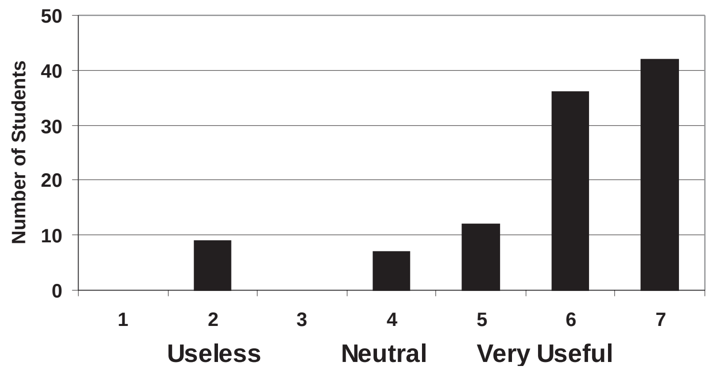

Figure 5. Subjective Evaluations by Trainees

more, anxiety did not decrease over trials ($p > 0.18$). This is especially important given the significant decrease in anxiety that occurs with IDCTT after only one trial [45]. Taken together, this suggests that DC-TRAIN 1.0 is psychologically realistic even after a number of trials. In other words, it creates and maintains a psychologically stressful environment that is necessary to prepare DCAs for what they would experience in an actual crisis [43].

Past evaluations of expert-based computerized simulations for Navy training have made use of various standards for success (such as, successful course completion [53]). By far, the most prevalent measures for all forms of computer aids in Navy instructional settings are those of user satisfaction [50]. The elicitation of user reactions is an ongoing part of the development of DC-Train. Most of the user evaluations are informal and qualitative and aimed at providing feedback to the developers so that they can fix "bugs" or make the simulator actions consistent with course material. Formal data on user reactions to the system, illustrated in Figure 5, show that DCA students rate DC-Train as over a 6.5 on a seven-point scale (anchored at 1 = completely useless and 7 = extremely useful). Although user satisfaction is necessary to continued use of training technologies, it is a contaminated and deficient measure of effectiveness. User reactions can reflect preexisting positive or negative biases about the technology, whereas they fail to capture aspects of performance critical to success in the context for which users are being trained. Thus, we turn to the true proof of the success of any training system: the performance of trainees.

In the first evaluation of DC-TRAIN, the low number of training trials per student (four trials at about 30 minutes each) was less than is desirable to affect expert performance. Another limitation was that the critiquing capability and scoring system were not yet available, meaning it was not possible to measure damage control performance in its entirety. Nevertheless, the resulting data, illustrated in Figure 6, show that firefighting errors were starting to decline by the fourth trial. Those students with three trials of practice on DC-TRAIN 1.0 made fewer firefighting errors on average on their IDCTT trial than students with less or no practice on DC-TRAIN 1.0.

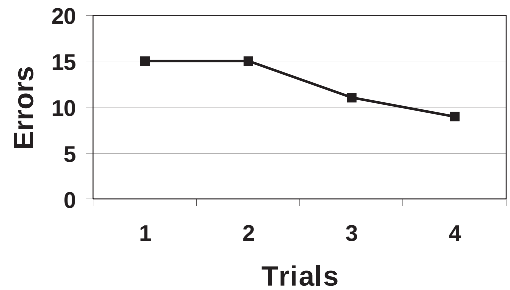

Figure 6. Error Scores Over Practice Trials

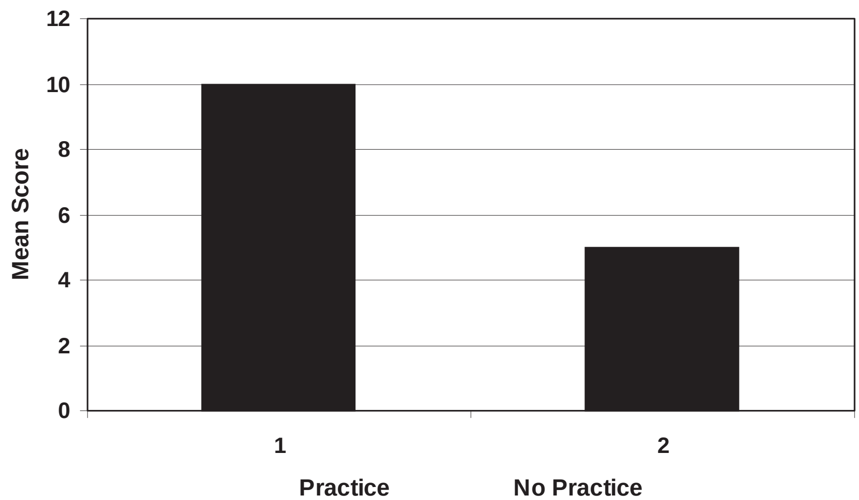

Figure 7. Average of Student Scores Across Scenarios

A second evaluation study with 19 DCA candidates was conducted to evaluate the effect of practice on performance. Nine DCA candidates received three practice trials using DC-Train, and then attempted to solve the criterion trial. Each trial used a different scenario (the criterion trial involved a different scenario than any of the three training trials). The other ten candidates did not receive practice. After controlling for military experience, the group receiving practice had significantly higher scores on average on the criterion scenario than the group who did not receive practice ($p < 0.05$). (Means are shown in Figure 7.) Eight of the nine trainees improved, and overall, practice accounted for 37 percent of the variance in performance scores. In other words, practicing on one set of scenarios was followed by improved performance on a different scenario. Because the criterion trial involved a different scenario than the practice trials, this improvement cannot have been due to participants merely learning how to beat a particular scenario. These data, and those of the first evaluation study, provide empirical support of the notion that multiple novel scenarios are beneficial for transfer of training to a new crisis. Although promising, this evaluation work has been limited by small sample sizes, scenarios restricted to events implemented in the most recent version of DC-Train, and low numbers of practice trials. Nevertheless, the ongoing evaluation of the system as well as human performance with the system is extremely valuable during system development and implementation, and provides a model for evaluation of other training systems for crisis decision-making.

Summary

CRISIS MANAGEMENT IS A COMPLEX SKILL that poses unique training problems. In this paper we have drawn on diverse literatures to analyze the issues involved in developing a crisis decision-making training program. We present an integrated framework for understanding these issues, and the advantages of computer-based systems with an immersive interface, expert critiquing, and scenario simulator. This framework should be helpful in evaluating and comparing various programs for crisis management training. In addition, we offer an example of such a trainer in the domain of ship damage control, DC-TRAIN, and describe its design and simulation and critiquing functions. We discuss the importance of empirical evaluations of training effectiveness using multiple criteria, and report data from experiments indicating that DC-TRAIN has realism and improves decision-making performance. These results are consistent with the argument that multiple novel scenarios have special value for training in crisis decision-making. Our approach to the assessment of realism can be generalized to the evaluation of other crisis training programs, whereas the automated assessment of human performance by the expert critiquing systems offers a useful model for training programs in general.

Discussion

ALTHOUGH HIGH-LEVEL EFFECTIVE HUMAN DECISION-MAKING during crises is vital in a vast array of situations, it is difficult to prepare personnel adequately. Crises pose serious barriers to both training and research on crisis management and training. Crises are, by definition, rare. Even when they do occur, the first priority is to handle them quickly to gain control of the situation and to minimize loss, not to train novices or conduct research [17]. The nonconventional approach of using systems such as DC-TRAIN that have an immersive interface, crisis simulation, and critiquing system have no such barriers. These systems provide the opportunity for immersive practice so trainees can develop the skills necessary for dealing effectively and efficiently with the whole-task decision-making while experiencing the overload and anxiety associated with real-life crises. Their unique capabilities address many training issues that are particularly problematic in training for crisis decision-making.

Simultaneously, crisis simulations create the opportunity for data collection under conditions reflecting the majority of a crisis' characteristics. This is enormously beneficial in evaluation research, especially on transfer of training. Perhaps even more significant is the potential for such sophisticated simulations to be used in basic research. Especially promising is the ability to conduct research on human decision-making performance under acute stress given that such research is critical to the success of crisis management training, but rarely feasible. The DC-TRAIN system can be used for research on the fundamental psychological process of crisis decision-making as well as for research in peripheral areas such as natural language critiquing and time perception. Through such research we can expect improvements in the theory behind the human performance issues of crisis management, which aid future development of training systems.

This paper argues that having a computer tutoring system is necessary but not sufficient for any crisis management training program. A complete training program in the crisis management domain requires the integration of three components: an immersive interface that creates a psychologically realistic experience; a simulator with numerous scenarios of varying attributes and difficulties; and a critiquing system that can provide immediate, specific, and accurate feedback while maintaining flexible communication between the system and user. DC-Train is a pioneering exemplar of such a system. It provides the foundation for further research for the development and creation of new computer-based training programs suitable to this domain. Yet, the design of any training systems demands careful consideration of the training issues discussed in this paper. An additional requirement for a successful system is a close matching between training needs in the context of interest and system capabilities. The basic characteristics of the DC-TRAIN system can be expected to generalize to training in many crisis domains; however, the specific problems of a particular domain may require new variations on a training solution. The framework provided can be a useful instrument to guide examination of these problems, and formulation of solutions.

Acknowledgments: This research was supported in part by ONR grant N00014-95-1-0749 and NRL contract N00014-97-C-2061. The authors gratefully acknowledge Vadim Bulitko, Scott Borton, Brian Weaver, Eugene Grois, and Ron Carbonari for assistance in data collection; Reeshad Dalal and Alyssa Mitchell for comments on earlier drafts; and thank the instructors, students, and command of SWOS for their support.

References

1. Ahituv, N.; Igbaria, M.; and Sella, A. The effects of time pressure and completeness of information on decision making. *Journal of Management Information Systems, 15,* 2 (Fall 1998), 153–172.

2. Balzer, W.K.; Doherty, M.E.; and O'Connor, R. Effects of cognitive feedback on performance. *Psychological Bulletin, 106,* 3 (1989), 410–433.

3. Balzer, W.K.; Sulsky, L.M.; Hammer, L.B.; and Sumner, K.E. Task information, cognitive information, or functional validity information: Which components of cognitive feedback affect performance? *Organizational Behavior & Human Decision Processes, 53,* 1 (1992), 35–54.

4. Baumann, M.R.; Sniezek, J.A.; and Buerkle, C.A. Self-evaluation, stress, and performance: A model of decision making under acute stress. In E. Salas and G. Klein (eds.), *Linking Expertise and Naturalistic Decision Making.* Mahwah, NJ: Lawrence Erlbaum, 2001, pp. 141–160.

5. Baumann, M.R.; Donovan, M.; Sniezek, J.A.; and Wilkins, D.C. Training effectiveness of an immersive multimedia trainer for acute stress domains: Ship damage control. Technical Report UIUC-BI-KBS-96011, Knowledge Based Systems Group, University of Illinois <<location>>, May 1996.

6. Benbasat, I., and Dexter, A.C. An investigation of the effectiveness of color and graphical information presentation under varying time constraints. *MIS Quarterly, 10,* 1 (1986), 51–80.

7. Bulitko, V.V. MINERVA-5—A multifunctional dynamic expert system. Masters dissertation, Department of Computer Science, University of Illinois, <<location>>, 1998.

8. Bulitko, V.V. Envisionment-based scheduling using time interval Petri networks: Representation, inference, and learning. Ph.D. dissertation, Department of Computer Science, University of Illinois, <<location>>, 2000.

9. Bulitko, V.V., and Wilkins, D.C. Automated instructor assistant for ship damage control. In <<editors>>, *Proceedings of the Eleventh Conference on Innovative Applications of Artificial Intelligence.* Conference held in Orlando, FL, July 18–22, <<publisher / location>>, 1999, pp. 778–785.

10. Camerer, C.F., and Johnson, E.J. The process-performance paradox in expert judgment: How can experts know so much and predict so badly? In K.A. Ericsson and J. Smith (eds.), *Toward a General Theory of Expertise: Prospects and Limits.* Cambridge: Cambridge University Press, 1991, <<page range>>.

11. Chi, M.; de Leeuw, N.; Chiu, M.H.; and LaVancher, C. Eliciting self-explanations improves understanding. *Cognitive Science, 18,* <<issue>> (1994), 439–477.

12. Cook, R.I., and Woods, D.D. Operating at the sharp end: The complexity of human error. In M.S. Bogner (ed.), *Human Error in Medicine*. Hillsdale, NJ: Lawrence Erlbaum, 1994, pp. 255–310.

13. Coovert, M.D.; Craiger, J.P.; and Cannon-Bowers, J.A. Innovations in modeling and simulating team performance: Implications for decision making. In Guzzo, Salas, & Associates (eds.), *Team Effectiveness and Decision Making in Organizations.* San Francisco: Jossey-Bass, 1995, <<page range>>.

14. Dennis, K.A., and Harris, D. Computer-based simulation as an adjunct to ab initio flight training. *The International Journal of Aviation Psychology, 8,* 3 (1998), 261–276.

15. Dixon, M.C., and Burns, J.L. Crisis theory, active learning, and the training of telephone crisis volunteers. *Journal of Community Psychology, 2,* 2 (1974), 120–125.

16. Dodani, M. The art of skill building without teaching. *The Journal of Object-Oriented Programming, 12,* <<issue>> (1999), 49–50.

17. Donovan, M.A.; Sniezek, J.A.; and Baumann, M. *Learning from Unusual Events: Decision Making in Crises.* Symposium sponsored by Managerial and Organizational Cognition and Organizational Behavior divisions of the Academy of Management, <<is this a paper presented? or the name of the symposium? what is the title of the paper / what is the name of the symposium>>, Cincinnati, OH, August 1996.

18. Driskell, J.E., and Salas, E. *Stress and Human Performance*. Mahwah, NJ: Lawrence Erlbaum, 1996.

19. Duchastel, P. Towards methodologies for building knowledge-based instructional systems. *Instructional Science, 20,* <<issue>> (1991), 349–358.

20. Edelson, D.C. Learning from cases and questions: The Socratic case-based teaching architecture. *Journal of the Learning Sciences, 5,* 4 (1996), 357–410.

21. Edland, A., and Svenson, O. Judgment and decision making under time pressure. In O. Svenson and A.J. Maule (eds.), *Time Pressure and Stress in Human Judgment and Decision Making.* New York: Plenum Press, 1993, pp. 27–40.

22. Einhorn, H.J., and Hogarth, R.M. Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology, 32,* <<issue>> (1981), 53–88.

23. Flin, R.; Slaven, G.; and Stewart, K. Emergency decision making in the offshore oil and gas industry. *Human Factors, 38,* 2 (1996), 262–277.

24. Gohm, C.L.; Baumann, M.R.; and Sniezek, J.A. Personality in extreme situations: Thinking (or not) under acute stress. *Journal of Research in Personality,* (in press).

25. Goldstein, I.L., & Associates. *Training and Development in Organizations.* San Francisco: Jossey-Bass, 1989.

26. Graesser, A.C., and Person, N.K. Question asking during tutoring. *American Educational Research Journal, 31,* <<issue>> (1994), 104–137.

27. Grise, M.L., and Gallupe, R.B. Information overload: Addressing the productivity paradox in face-to-face electronic meetings. *Journal of Management Information Systems, 16,* 3 (Winter 2000), 157–185.

28. Grois, E.; Hsu, W.H.; Voloshin, M.; and Wilkins, D.C. Bayesian network models for automatic generation of crisis management training scenarios. In <<editors>>, *Proceedings of the Ninth Conference on Innovative Applications of Artificial Intelligence, IAAI-98.* Conference held in Madison, WI, <<publisher / locations>>, July 1998, pp. 1113–1120.

29. Hart, S.G., and Staveland, L.E. Development of NASA-TLX (task load index): Results and empirical and theoretical research. In P.A. Hancock and <<initial>> Meshkati (eds.), *Human Mental Workload.* Amsterdam: North Holland, 1988.

30. Hays, R.T., and Singer, M.J. *Simulation Fidelity in Training System Design*. New York: Springer-Verlag, 1989.

31. Hume, G.; Michael, J.A.; Rovick, A.A.; and Evens, M. Hinting as a tactic in one-on-one tutoring. *Journal of the Learning Sciences, 5,* 1, (1996), 23–47.

32. Kantowitz, B.H. Selecting measures for human factors research. *Human Factors, 34,* 4 (1992), 387–398.

33. Kluger, A.N., and DeNisi, A. Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119,* 2 (1996), 254–284.

34. Koschmann, T.D.; Myers, A.C.; Feltovich, P.J.; and Barrows, H.S. Using technology to assist in realizing effective learning and instruction: A principle approach to the use of computers in collaborative learning. *The Journal of the Learning Sciences, 3,* 3 (1994), 227–264.

35. Majchrzak, A. *The Human Side of Factory Automation.* San Francisco: Jossey-Bass, 1988.

36. Mao, J.Y., and Benbasat, I. The use of explanations in knowledge-based systems: Cognitive perspectives and a process-tracing analysis. *Journal of Management Information Systems, 17,* 2 (Fall 2000), 153–179.

37. McKinney, E.H. Flight leads and crisis decision-making. *Aviation, Space and Environmental Medicine, 64,* 5 (1993), 359–362.

38. Mouloua, M., and Koonce, J.M. *Human-Automation Interaction*. Mahwah, NJ: Lawrence Erlbaum, 1997.

39. Orasanu, J.M. Decision-making in the cockpit. In E. Wiener (ed.), *Cockpit Resource Management.* San Diego: Academic Press, 1993, pp. 137–172.

40. Quanjel, M.M.; Willems, A.J.; and Talen, A.N. CRISISLAB: Evaluation and improvement of crisis management through simulation/gaming. *Simulation & Gaming, 29,* 4 (1998), 450–455.

41. Salas, E.; Bowers, C.A.; and Rhodenizer, L. It is not how much you have but how you use it: Towards a rational use of simulation to support aviation training. *The International Journal of Aviation Psychology, 8,* 3 (1998), 197–208.

42. Salas, E.; Driskell, J.E.; and Hughes, S. Introduction: The study of stress and human performance. In J.E. Driskell and E. Salas (eds.), *Stress and Human Performance.* Mahwah, NJ: Lawrence Erlbaum, 1996, pp. 1–45.

43. Schrah, G.E.; Chernyshenko, O.S.; Sniezek, J.A.; Baumann, M.R.; Bulitko, V.; Wadlington, P.; Wilkins, D.C.; and Borton, S. Evaluation of DC-TRAIN: Performance measurement and validation. Investigating reliability and validity of the DC-TRAIN scoring system: Two field studies. Technical report. Office of Naval Research, <<location>>, 2000.

44. Shou, G.; Wilkins, D.C.; Hoemann, M.; Mueller, C.; Tatem, P.A.; and Williams, F.W. Supervisory control system for ship damage control: Volume 2—Scenario generation and physical ship simulation of fire, smoke, flooding and rupture. Report NRL/MR/6180-01-8572, Naval Research Laboratory, Washington, DC, August 24, 2001.

45. Sleeman, D., and Brown, J.S. *Intelligent Tutoring Systems.* Orlando, FL: Academic Press, 1982.

46. Sniezek, J.A., and Reeves, A. Aiding decision making with aggregate information in a laboratory study. In <<editors>>, *Proceedings of the Hawaii International Conference on Systems Sciences,* <<publisher / location>>, January 1984, pp. 603–608.

47. Sniezek, J.A., and van Swol, L. Trust, confidence, and expertise in Judge Advisor Systems. *Organizational Behavior and Human Decision Processes, 84,* 2 (2001), 288–307.

48. Spielberger, C.D.; Gorsuch, R.L.; and Lushene, R.E. *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologist Press, 1970.

49. Spielberger, C.D.; Gorsuch, R.L.; Lushene, R.; Vagg, P.R.; and Jacobs, G.A. *State-Trait Anxiety Inventory for Adults: Sampler Set, Manual, Test, Scoring Key.* Palo Alto, CA: Mind Garden, 1983.

50. Sticht, T.; Ellis, J.; Montague, W.; Quellmalz, E.; and Slappy, J. Combining environmental design and computer programs to enhance learning in Navy technical training. *Military Psychology, 5,* 1 (1993), 63–75.

51. Su, Y., and Lin, D.M. The impact of expert-system-based training on calibration in emergency management. *Computers in Human Behavior, 14,* 1 (1998), 181–194.

52. Thompson, L., and DeHarpport, T. Social judgment, feedback, and interpersonal learning in negotiation. *Organizational Behavior & Human Decision Processes, 58,* 3 (1994), 327–345.

53. Van Matre, N., and Robinson, C.R. The automated maneuvering board training system: An in progress implementation evaluation. *Journal of Computer-Based Instruction, 13,* 2 (1986), 35–38.

54. Wilkenfeld, J.; Kraus, S.; and Holley, K.M. The negotiation training model. *Simulation & Gaming, 29,* 1 (1998), 31–43.

55. Ye, L.R., and Johnson, P.E. The impact of explanation facilities on user acceptance of expert systems advice. *MIS Quarterly, 19,* 2 (1995), 157–172.

56. Zakay, D., and Wooler, S. Time pressure, training and decision effectiveness. *Ergonomics, 27,* 3 (1984), 273–284.