Special Report 35

1997 IN-HOUSE RESEARCHER COLLOQUIUM

February 1998

United States Army Research Institute
For the Behavioral and Social Sciences

FOREWORD

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) held its first formal In-house Researcher Colloquium on 20 November 1997 in Alexandria, Virginia. The eight researchers who presented research findings at the colloquium represented ARI's Armored Forces Research Unit, the Automated Training Methods Research Unit, the Fort Leavenworth Research Unit, the Infantry Forces Research Unit, the Organization and Personnel Resources Research Unit, the Rotary Wing Aviation Research Unit, the Selection and Assignment Research Unit, and the Simulator Systems Research Unit. Each research topic (and researcher) was specifically selected by the Research Unit Chief as an example of the best of research being performed at the unit.

The main purpose of the in-house colloquium was to provide a structured opportunity for discussion among ARI's more junior researchers, across units, who represented the broad areas of U.S. Army behavioral and social science research conducted by ARI. This monograph provides brief summaries of the research and biographies of the researchers. The monograph also serves as an example of the range of research topics being addressed by in-house researchers at ARI as well as of the backgrounds of ARI's research staff. Because of the success of the colloquium, we expect it to be an annual event, and we will look for additional opportunities to increase cross-unit communication.

EDGAR M. JOHNSON
Director

Table of Contents

PEER EVALUATIONS: WHAT DO THEY BUY YOU?

Michelle M. Zazanis
Organization and Personnel Resources Research Unit
Alexandria, VA

Abstract

Peer evaluations have historically shown high predictive validity, but the reason for this strength has not been clear. This research demonstrated that interpersonal performance and effort are key dimensions of peer assessments that may account for their historical strength. Longitudinal research is needed to explicitly link these dimensions of peer evaluations with future performance.

Introduction

Peer evaluations have a long history of research, particularly in military settings (for reviews, see Downey & Duffy, 1978; Reilly & Chao, 1982; Kane & Lawler, 1978). Research suggests that they are extremely reliable (Hollander, 1956,1957; Gordon & Medland, 1964) and highly valid, successfully predicting factors such as how well students perform in training, their grades in the training course, whether they graduate, how quickly soldiers are promoted, and how well soldiers perform in combat situations (Downey & Duffy, 1978; Dugan, 1953). Studies that have compared peer evaluations to other methods of assessment have shown that peer evaluations are more highly predictive of future performance than measures such as military course grades (O'Connor & Berkshire, 1958), instructor ratings (Kraut, 1975; Wherry & Fryer, 1949), objective tests, and supervisory ratings (Williams & Leavitt, 1947).

The reason for their strong predictive validity, however, has not been clear. Different groups of raters (supervisors, peers, self, subordinates) have different views of an individual's performance and, therefore, use different information in making evaluations (Borman, 1974; Murphy & Cleveland, 1991; Knapp & Campbell, 1993). Mumford (1983) has suggested that peers are very focused on each other's performance in order to determine the quality of their own performance. This focus on social comparisons, he suggested, functions to increase the accuracy of their assessments of others' task performance skills.

In contrast to an individual's instructor or supervisor, peers typically have a greater number of opportunities to observe an individual's performance, and have a more interactive perspective of the individual. Murphy and Cleveland (1991) have suggested that the supervisor's presence generally elicits an individual's maximal performance, not typical performance. The peer's perspective, then, may provide better opportunities to gain insights into an individual's attitudes and personality characteristics. Peer evaluations may contain more information than supervisor evaluations about an individual's personality and interpersonal skills (Borman, White, & Dorsey, 1995; Oppler, Peterson, & McCloy, 1994). In addition, peers may be able to distinguish better than supervisors between effort and skill dimensions of performance (Klimoski & London, 1974). Thus, it may be that peer evaluations of performance demonstrate high predictive validity because they include more information about the interpersonal and motivational dimensions of performance than do supervisor ratings.

This research compared ratings of overall performance made by peers and supervisors, and investigated whether there were implicit differences between them in the importance attributed to task, interpersonal, and motivational dimensions. Peer raters were expected to give greater importance to interpersonal skills and motivation than supervisors.

Method

Subjects & Setting

Subjects were 239 enlisted male Army soldiers attending the first phase of training for Special Forces at the Special Forces Qualification Course (SFQC) between March and October of 1995. Soldiers attending this course were selected for training based on their performance during a 21-day assessment center (Special Forces Assessment and Selection). The first phase of SFQC teaches land navigation and small unit tactics and spans 25 days. Soldiers are assigned to squads of 10-14 members who function as a team during training. Each team is assigned a primary trainer, a Special Forces-qualified soldier who trains and evaluates student performance.

Measures

Ratings of students' task performance, interpersonal performance, effort/persistence, overall performance, and predicted future performance on an SF A-team, were collected from peers and trainers at the end of Phase I. The rating format used a five-point Likert scale and provided the rater with behavioral indicators of performance.

Results

Ratings of overall performance and predicted future performance were regressed separately and simultaneously on ratings of task performance, interpersonal performance, and effort, for both trainers and peers. Results showed that each of the three dimensions explained unique variance in peer ratings of overall performance and future performance (see Table 1). For trainers, however, unique variance in overall current performance was explained by task performance and effort only; interpersonal performance explained significant unique variance only in the trainers' ratings of future performance. For both overall current performance and future performance, however, interpersonal performance ratings had a significantly stronger regression coefficient for peers than for trainers.

Table 1. Comparison of Regression Coefficients for Peer and Trainer Raters

1. Overall Current Performance regressed on Component Ratings:

	Peers		Trainers		Bp-Bt
	B	R2	B	R2	
Task	.64**		.70**		-.06
Interp	.17**		.09	.08*	
Effort	.22**	.95	.17**	.84	.05

2. Future Performance regressed on Component Ratings:

	Peers		Trainers		Bp-Bt
	B	R2	B	R2	
Task	.46**		.57**		-.11*
Interp	.28**		.16**		.12**
Effort	.30**	.93	.20**	.77	.10

Note: *p<.05 and **p<.01

As indicated in Table 2, when peers rated future performance, the regression coefficients for interpersonal performance and effort were significantly larger and that for task performance was significantly smaller. For trainers, coefficients for interpersonal performance and effort increased, but not significantly, and the coefficient for task performance significantly decreased.

Table 2. Differences in Regression Coefficients Using Current and Future Performance

1. Differences in Peer Predictor coefficients for each criterion:

	Criteria		
	Current Performance	Future Performance	Bf-Bc
Regressed on:	Bc	Bf	
1. Task	.64	.46	-.18**
2. Interp	.17	.28	.11**
3. Effort	.22	.30	.08*

2. Differences in Trainer Predictor coefficients for each criterion:

	Criteria		
	Current Performance	Future Performance	Bf-Bc
Regressed on:	Bc	Bf	
1. Task	.70	.57	-.13*
2. Interp	.09	.16	.07
3. Effort	.17	.20	.03

Note: *p<.05 and **p<.01

Discussion

These analyses were designed to examine the premise that peers include more information than trainers about interpersonal performance and effort in their ratings of an individual's overall current and future performance; this premise was supported. While both peers and trainers included information about task-specific performance and effort in their overall assessment of an individual's current performance, only peers included unique information about interpersonal performance in this assessment. In addition, when predicting future performance, while both peers and trainers significantly decreased the amount of unique information they included regarding task-specific performance, only peers significantly increased the amount of information they included concerning both interpersonal performance and effort.

This research suggests that interpersonal performance and effort are key factors in the structure of peer assessments that may account for their historical strength as assessment tools, especially in predicting future performance. Whether this extends to the predictive strength of peer evaluations, however, is something that cannot be directly answered by these data, but would require a longitudinal analysis that followed these individuals to their actual performance on-the-job. These results strongly suggest, however, that peer evaluations may offer a method of measuring motivation and interpersonal skills, attributes that are difficult to measure in other ways.

References

Borman, W.C. (1974). The rating of individuals in organizations: An alternate approach. *Organizational Behavior and Human Performance, 12,* 105-124.

Borman, W.C., White, L.A., & Dorsey, D.W. (1995). Effects of ratee task performance and interpersonal factors on supervisor and peer performance ratings. *Journal of Applied Psychology, 80(1),* 168-177.

Downey, R.G., & Duffy, P.J. (1978). *Review of Peer Evaluation Research.* (Technical Paper 342). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social .Sciences.

Dugan, R.D. (1953). Comparison of evaluation of B-29 crews in training and in combat. *American Psychologist, 8,* 343-344.

Gordon, L.V., & Medland, F.F. (1964). The cross group stability of peer ratings of leadership potential. *Personnel Psychology, 18,* 173-178.

Hollander, E.P. (1956). Interpersonal exposure time as a determinant of the predictive utility of peer ratings. *Psychological Reports, 2,* 445-448.

Hollander, E.P. (1957). The reliability of peer nominations under various conditions of administration. *Journal of Applied Psychology, 41,* 85-90.

Kane, J.S. & Lawler, E.E., III (1978). Methods of peer assessment. *Psychological Bulletin, 85,* 555-586.

Klimoski, R.J. & London, M. (1974). Role of the rater in performance appraisal. *Journal of Applied Psychology, 59*(4), 445-451.

Knapp, D.J. & Campbell, J.P. (1993). *Building a joint-service classification research roadmap: Criterion-related issues.* (AFMC AL/HR-TP-1993-0028). Brooks Air Force Base, TX: Human Resources Directorate Manpower and Personnel Research Division.

Kraut, A.I. (1975). Prediction of managerial success by peer and training staff ratings. *Journal of Applied Psychology, 60,* 14-19.

Mumford, M.M. (1983). Social comparison theory and the evaluation of peer evaluations: A review and some applied implications. *Personnel Psychology, 36(4),* 867-881.

Murphy, K.R. & Cleveland, J.N. (1991). *Performance Appraisal.* Boston, MA: Allyn and Bacon.

O'Connor, W.F., & Berkshire, J.R. (1958). Comparison of the pre-flight OLQ grade and the leadership peer rating as predictors of training failure (58-16). Naval Aviation Medical Center, Pensacola, FL: U.S. Naval School of Aviation Medicine.

Oppler, S.H., Peterson, N.G., & McCloy, R.A. (1994). *A comparison of peer supervisor ratings as criteria for the validation of predictors.* Paper presented at the 9th Annual meeting of the Society of Industrial and Organizational Psychology, Nashville, TN.

Reilly, R.R. & Chao, C.T. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology, 35,* 1-62.

Wherry, R.J., & Fryer, D.H. (1949). Buddy ratings: Popularity contest or leadership criteria? *Personnel Psychology, 2,* 147-159.

Williams, S.B., & Leavitt, H.J. (1947). Group opinion as a predictor of military leadership. Journal of Consulting Psychology, *11,* 283-291.

Dr. Zazanis holds a B.A. in psychology from the University of Virginia and an M.A. and Ph.D. in industrial/organizational psychology from George Mason University. She has been working with the Special Forces Team since 1991, when she came to ARI as a Consortium Research Fellow. In 1995, ARI hired her full time as a research psychologist. With the Special Forces Team, she has worked on a variety of research projects involving the selection and training of soldiers for Special Forces, including the development of a training program for selection program personnel, and the development of new peer evaluation systems and climate surveys for both the selection and training programs.

BATTLESTAFF TRAINING: A STRUCTURED APPROACH

Bruce Sterling
Armored Forces Research Unit
Fort Knox, KY

Abstract

Battle staff proficiency has a substantial effect on battle outcome. Research and insights from combat training centers demonstrate that staffs have significant problems in many areas of information management. In response to these problems, the Army Research Institute has developed a research and development program applying principles of structured training to improving battle staff performance. This paper provides an outline of structured training, and a prototype strategy for applying structured training to individual staff officer, small staff group and battle staff training. Future directions of this research are briefly discussed.

Introduction

The work of the battle staff is critical to the success of military operations. Research literature defines the staff as a decision making team that assists the commander in obtaining situation awareness, making decisions and implementing decisions (McIntyre and Salas, 1995). As such, the role of the staff is unique. Even though individuals, crews, platoons and company-teams are superbly trained, without direction they will have little, if any, effect in an operation.

Army Problem

Both empirical research and observations by subject matter experts suggest problems with battle staff performance. Thompson, Pleban and Valentine (1994) reviewed five studies documenting performance of battalion level staffs at Combat Training Centers (CTCs). They found weaknesses by both staff position and function. Institutional training is often provided too late to help staff officers in their jobs. Further, staffs have little stability over time. Recent trends reported by Observer Controllers (OCs) at CTCs (e.g.; Department of the Army, 1997) provide more specific information on staff deficiencies. These trends include problems in battle tracking, difficulties in integrating the information received and problems in disseminating the information to other staff sections. Also, OCs state that staffs do not consistently analyze the information acquired, thus failing to provide the commander with situation awareness, predictive analyses, and recommendations for future actions. Thus problems with information management, from obtaining and recording information, through integration and communication to analysis, appear to be rife within the staff.

Approach

Military researchers have developed a training paradigm to train staffs. Researchers with the U.S. Navy refer to the technique as event based training or guided self practice (Johnston, Smith-Jentsch, and Cannon-Bowers, 1997), while U.S. Army researchers call the method structured training (Burnside, Leppert & Myers, 1996). The technique includes four phases. First, the tasks, processes or behavioral dimensions to be trained are determined. An example of a task would be "track the battle." An example of a process or behavioral dimension might be "analyze information to determine its significance" or "adaptability and flexibility", respectively. Next, a realistic scenario (usually run in some sort of simulation) is designed in which events occur which cause these tasks, processes or dimensions to be performed. Then, behavioral measures of performance on these tasks, processes or dimensions are developed for use by observers. Finally, the observers use these behavioral measures to collect information to be used to provide feedback to the staff. The feedback is provided by observers after the exercise, using open-ended questions designed to cause participation by staff members. Staffs then decide on which tasks, processes or dimensions they performed adequately and on which they need to improve, choosing a goal for the next exercise.

Research & Development Program

The U.S. Army Research Institute is engaged in research and development for a variety of staff training programs designed to provide structured training for individual staff members, small staff groups and entire staffs in critical staff tasks and processes. The battle staff training system (BSTS) provides new staff officers with a course in their staff position at their duty station (Andre & Salter, 1995). The course consists of computer assisted training designed to train staff officers at battalion and brigade level in individual staff skills. It is based on the principle of test-train-test. Each course is composed of subjects and each subject is composed of lessons. Material includes both paper and computer based instruction. Officers having difficulty with certain courses can access remedial training developed to correct the most commonly detected shortfalls. Comprehensive assessment tests, designed to measure ability to use, versus merely recall, information learned in the courses have also been developed.

The Staff Group Trainer (SGT) provides computer assisted training for small battalion and brigade staff groups (e.g.; intelligence section, operations section, fire support section) in staff processes (e.g., analyze, communicate, recommend) during the execution phase of battle (BDM Federal, 1995). All message input from subordinate elements, adjacent units and higher headquarters is scripted. Staff sections receive the messages at computer workstations and must post information to their situation maps, extract information for the purpose of coordination with other staff members, consolidation, reporting and making recommendations to the commander or higher headquarters. Observers are provided with observer checklists describing staff behaviors to look for in association with specific key messages (events). These checklists assist the observers in providing feedback during After Action Reviews. Both battalion and brigade training support packages are divided into three tables: staff section, command post and command and control. The staff section table is divided into separate modules for each staff section. These modules contain exercises training information processing within the staff section. The command post table contains separate modules for each command post. These modules contain exercises training communicating, coordinating and integrating information among staff sections. The command and control table contains modules for different missions (i.e., movement to contact, attack, defend). Each of these modules contain exercises training staff processes taking place among command posts such as disseminating orders or directing units.

The Combined Arms Operations at the Brigade Level, Realistically Achieved through Simulation (COBRAS) research and development effort (Graves, Campbell, Leibrecht, Deter, Hoffman, Ford and Campbell, In preparation) involves live and constructive structured simulation based training for staffs. It trains brigade staffs in the plan and prepare, consolidation and reorganization stages as well as the execute stage. The COBRAS program uses brigade staff vignettes to train selected small staff groups integration and synchronization skills. Brigade staff groups are given an introduction to COBRAS, including a discussion of the tasks, conditions and standards to which they will be trained. Then the staff is given a scenario of how the political and military situation developed (road to war), an abbreviated corps order (mainly commander's intent), and a detailed division order, with appropriate annexes, graphics and other materials. An observer team plays the role of higher headquarters. The brigade staff small groups then develop and execute parts of the brigade order and related products. The observer team conducts after action reviews at certain logical points during planning, preparation and execution, providing comments on the processes related to staff integration. Virtual staff training in mission execution is then provided in JANUS (Burnside, Leppert, & Meyers, 1996). This training focus on the principle staff members only and uses role players to represent subordinate and higher staff members.

Training in the plan, prepare and execute phases for the entire brigade staff is provided in COBRAS brigade staff exercise. The training support package provided is much like that described above for COBRAS vignettes. The observer team is also responsible for ensuring the exercise trains the intended tasks and for providing an after action review of the staff. Integrated exercises have been developed for movement to contact, area defense and deliberate attack missions using the NTC data base.

In the final phase of the overall R&D effort, subordinate (battalion) staffs are involved in planning, preparation and execution of combat operations. For brigades, this training is currently being developed. One possibility is a synthetic theater of war (STOW). In one variant of STOW, subordinate leaders of one battalion are trained in virtual simulation; and staff are trained in a constructive simulation for all battalions. The brigade staff is also in a constructive environment. This training follows the same principles and design outlined above, except that subordinate commanders come from the training unit and the materials provide them little external guidance outside of the OPORD. That is, subordinate commanders may provide no (or incorrect) information about battlefield events, and may incorrectly execute orders from brigade staff. In other words, these exercises introduce more "fog of war" for the staff to work through.

Future research and development will focus on modifying structured training programs to reflect digital systems, including tasks that are performed differently in digital systems (e.g., abbreviated troop leading procedures). We will also attempt to make use of automated feedback to reduce the overhead (e.g., interactors and observers) required to run these programs.

In summary, ARI is attempting to meet an Army need through research based structured training, adopting the technique to provide individual, small group and collective staff training for the future Army.

References

Andre, C. R. & Salter, M. S. (1995). Battalion battle staff training system. Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences.

BDM Federal (1996b). Commander/Staff Trainer (C/ST) project design. Fort Knox, KY: BDM Federal Incorporated.

Burnside, B. L., Leppert, M. A. & Meyers, W. E. (1996). Virtual training program orientation guide (Research Product 97-02). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Department of the Army (1997). Combat Training Center (CTC) trends (97-6): Joint Readiness Training Center 4QFY96 & 1QFY97. Fort Leavenworth, KS: Center for Army Lessons Learned.

Graves, C. R., Campbell, C. H., Leibrecht, B. C., Deter, D. E., Hoffman, R. G., Ford, L. A. & Campbell, R. C. (In preparation). Combined arms operations at brigade level, realistically achieved through simulation (COBRAS) Project: History and lessons learned (Draft Final Report). Alexandria VA: Human Resource Research Organization.

Johnston, J. H., Smith-Jentsch, K. A. & Cannon-Bowers, J. A. (1997). Performance measurement tools for enhancing team decision making training. In M. T. Brannick, E. Salas & C. Prince (Eds.), Team performance assessment and measurement: Theory, methods, and applications (pp 311-330). Mahway, NJ: Lawrence Erlbaum.

McIntyre, R. M. & Salas, E. (1995). Measuring and managing for team performance: Emerging principles from complex environments. In R. A. Guzzo & E. Salas (Eds) Team effectiveness and decision making in organizations. San Francisco: Josey-Bass Publishers (pp 9-45).

Thompson, T. J.; Pleban, R. J.; & Valentine, P. J. (1994). Battle staff training and synchronization: The integration of functions critical to combat operations. In R. F. Holz, J. H. Hiller & H. H. McFann (Eds), Determinants of effective unit performance. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences (pp 179-230).

Dr. Sterling received his Ph.D. in Social Psychology from the University of Delaware in 1977. He has worked as a research psychologist and operations researcher for ARI, US Army Soldier Support Center, TRADOC Analysis Command-White Sands Missile Range (TRAC-WSMR) and US Army Europe (USAREUR). He has performed research in the areas of personnel and training.

DEVELOPING EXPERTISE IN BATTLEFIELD THINKING

James W. Lussier
Fort Leavenworth Research Unit

Abstract

The Fort Leavenworth Research Unit (FLRU) of the U.S. Army Research Institute is beginning work on a new program called Battle Command Skills: Preparing for the Future Battlefield. The organizing concept is to explicitly identify and train Army leaders in the thinking skills they need for planning and executing military operations. The effort employs training methods used by the Soviets in developing expertise in chess champions. These methods use a psychological, skill-building approach rather than relying on developing a purely intellectual understanding of the field. The report describes the plan of developing automated training in battlefield thinking skills. A variety of recent FLRU projects in battle command, knowledge elicitation, mental models, problem solving, tactical planning, and practical thinking, and command and control prepare the research unit to implement its plan. Some of the issues to be resolved involve whether to train for novel or typical situations, under current or future conditions, and to what extent to employ procedural or non-procedural means of changing behavior.

Introduction

The Fort Leavenworth Research Unit of the Army Research Institute is beginning work on a new program called Battle Command Skills: Preparing for the Future Battlefield. The organizing concept is to explicitly identify and train Army leaders in the thinking skills they need for planning and executing military operations. The spirit of the effort is well illustrated by recent events in the world of chess.

Soviets Stun the Chess World

Traditional methods of developing expertise in chess are similar to those of developing military acumen. First, one must study theory and doctrine. As an example from chess, opening theory comprises well-studied ways of beginning the game that have developed over centuries of research. Other theoretical concepts from chess are center control, space advantage, the initiative, and piece development. Major tenets of chess theory comprise time, space, and force which can be compared with the tenets of army operation: initiative, agility, depth, synchronization, and versatility. Military art is also based on specific doctrine, which, for the U.S. Army is capstoned by FM 100-5, Operations. Another method of developing expertise is to study the actions of great exponents of the craft. It is as typical in chess to study great games from the past, dissecting them in detail, as it is to examine historical battles and campaigns. Both fields contain proceduralized knowledge for application in typically occurring situations. An example from chess would be exploitation of the advantage of a knight over a bishop in blocked positions. There is a sequence of steps, which compose an overall plan in these situations. Similarly many military situations have analogous model procedures, such as the steps in breaching a minefield which is defended by an enemy unit. In both fields, study is insufficient without practice. The development of expertise requires experience, which is typically obtained in both low-stakes, practice situations, such as casual chess games or battle simulations, and real experience obtained by playing in tournaments or fighting a war. These methods are time-honored. Indisputably they are requirements on the road to mastery in either field.

For decades the Soviet Union has dominated the international chess scene. Since 1927, the world champion has been from the Soviet Union (although Alekhine defected at the revolution and played under the French flag) except for the years 1935-1937, when the title was held by Euwe, from Holland, and the years from 1972 to 1975 when Bobby Fischer was the World Champion. Also the vast majority of title matches since 1927 have been played by Soviets against Soviets. Furthermore, this success at the top does not capture the amazing depth of the Soviet chess machine. A USSR against the Rest of the World match was held in which the Soviets triumphed. It was generally assumed that the Soviet success was due to an especial effort made at identifying and supporting promising players. It was not imagined that the Soviets followed any different methods of developing experts than those traditional methods discussed above. With the breakup of the USSR, chess academies, hitherto secret, became capitalistic enterprises, publishing books describing their methods. The chess world was stunned to see that the Soviets, while not discounting the traditional approaches, incorporated an entirely new method. They did not merely study the game of chess. They studied themselves as players of the game.

Soviet trainers viewed the game as a sport and a struggle, rather than as an intellectual endeavor. They identified specific cognitive tasks that the player would have to perform, for example, searching a problem space mentally, allocating planning time, visualizing future positions, having creative ideas, and evaluating positions accurately in a time efficient manner. These behaviors were specifically trained, as were modes of thought, which the Soviets termed schematic thinking and prophylactic thinking. Players were taught how to control emotionally- induced errors which might occur when they were surprised by an unanticipated move, or when they were defending a difficult, even losing, position. They were taught to develop an intuitive feel for positions so they could sense when caution was required, when a bold decisive move had to be played, or when slow maneuvering was advantageous. They were taught methods to induce the opponent to make errors, when not to hurry in progressing the flow of the game, and how to play "cat-and-mouse" with a threat. They were taught to continually put themselves in their opponent's mind, to view the game from that perspective and to strive to understand the opposing thinking and the rationale motivating each move. They took a hard, critical look at their own thinking processes, ferreting out their weaknesses and targeting them. This only begins to describe the revolutionary approach to the game taken by the Soviet School; everything they do and write is suffused with a psychological mindset. And they accomplish all this while still recognizing vast stylistic differences among the approaches to the game of various practitioners.

A New Approach to Battlefield Thinking

In retrospect, none of the Soviet innovations seem extraordinary. One feels they should have been evident. It is only necessary for the serious chess player to begin to read such works as *Think Like a Grandmaster* by Alexander Kotov, or *Positional Play* by Mark Dvoretsky and Artur Yusapov (a preeminent trainer and his star pupil) to be taken with the realization that this is something entirely different. The books are packed with exercises to actively involve the reader. Traditional instruction at its best left the reader feeling like a spectator to a lecture. The US Army provides solid support for the traditional requirements of developing expertise with frequently updated doctrinal publications, a career-long progression of institutional education, and a great investment in simulation-based exercise capability. However, the development of thinking skills is left to occur naturally. Where the Army development process does explicitly address thinking skills, they do so in a weak, simplistic, outmoded, and ineffective manner, e.g., teaching symbolic logic syllogisms to promote better reasoning. Procedures have been developed to structure some cognitive tasks such as tactical planning. Such structure can be useful, especially for group tasks, but the performance issues are not addressed. For example, the step of commander's guidance is included in the deliberate decision making process. Commanders are not, however, explicitly trained on how to discern the critical elements of a situation and communicate it to subordinates. Such skills are allowed to develop naturally in the context of traditional methods, e.g., academic study and training exercises. Nor are planners taught how to critically evaluate options. Instead they are given a quantitative method which is completely inappropriate in the great majority of battlefield situations in which the officers will find themselves.

Our intent at the Fort Leavenworth Research Unit is to develop and deliver training in thinking skills which derives, in spirit, from the work of Soviet chess trainers. We will continue our research to understand the battlefield thinking skills underlying such cognitive tasks as situation assessment, visualization, planning, and dynamic decision making. We envision delivering automated training that is self-administered and paced and used by army officers as part of their self-development program. Much of what is planned will be made available to officers using internet technology. The effort to develop automated instruction at the Command and General Staff College (CGSC) is also a targeted delivery vehicle.

Building Blocks for Success

The Research Unit has a long history of relevant research and development efforts, which will be instrumental in developing the new concepts in training battlefield thinking. We studied expertise and discovered key differences in what experts and novices think about in a tactical planning task (Deckert et al, 1996). Participating in the Battle Command Battle Laboratory's focused rotations on battle command allowed development of key concepts and identification of critical factors in the art of battle command (Lussier & Saxon, 1994). How tactical thinkers organize and speak about their knowledge was the subject of a study by Michel (in preparation). In addition to these specific elements of battle command expertise we have examined generalized thinking strategies. Pounds and Fallesen (in preparation) surveyed the problem solving literature, identified 66 strategies, and studied the occurrence of the strategies in tactical problem solving situations. Members of the Research Unit wrote and presented a course in practical thinking skills at the CGSC, including such topics as taking multiple perspectives, recognizing assumptions, metacognition, and integrative thinking (Fallesen et al, 1995). We have also examined much more specific tactical skills, devising a measure of the ability of an officer to visualize events from tactical radio reports and to make accurate predictions in battalion level armor exercises (Solick et al, 1997). The paradigm developed in the visualization study was combined with much of the work mentioned above to develop the concept for an automated training program (Lussier, Michel, & Frame, 1997) which could serve as a focal point in the training of battlefield thinking skills. Finally, two recent efforts are worth mentioning. The first is the development of software to build a workstation to link experimenter and subject and support a wide variety of knowledge elicitation and mental model research (Zacharias et al, 1997). The workstation should also strongly support the development and testing of automated training materials. The second effort involves collecting stories from experienced officers describing how they learned a valuable lesson in battle command. Not only will the collection deepen our insight into battle command, it will provide an excellent vehicle for transmitting these surrogate lessons.

A host of issues remain to be answered, but that is expected with an effort of this nature. One of the most evident issues is whether to provide training in general or specific skills, i.e., to train thinking skills for use in battlefield situations or to train battlefield thinking skills. A related issue involves whether to emphasize performance in novel or in typical situations. Another issue is how to change thinking behavior without prescribing sets of procedures, or if such sets prove necessary how to promote their internalization so that they do not disrupt thought. Finally, we must address the introduction of new technology into the system, a process that has marked effect on the cognitive tasks required of the military officer. The project represents a number of strong challenges, however, the effort is worthwhile as it directly impacts ARI's mission, building the ultimate smart weapon - the American soldier.

References

Deckert, James C., Entin, Elleen B., Entin, Elliot E., MacMillan, Jean (1996). *Military Command Decisionmaking Expertise.* (ARI Research Note 96-16). Alexandria VA. US Army Research Institute for the Behavioral and Social Sciences. (AD A306 801)

Dvoretsky, M. & Yusupov A. (1996). *Positional Play.* London: Batsford.

Fallesen, J.J., Michel, R.R., Lussier, J.W. & Pounds, J.C. (1995) *Practical Thinking, Innovation in Battle Command Instruction.* (ARI Technical Report 1037). Alexandria VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A310 096)

Kotov, Alexander. (1978). *Think Like A Grandmaster.* London: Batsford.

Lussier, J.W., Michel, R., & Frame, A. (1997). NTC-CD System: Recreating the NTC Experience (ARI Study Report 95-01). Alexandria VA. US Army Research Institute for the Behavioral and Social Sciences. (AD A328 363)

Lussier, J.W. & Saxon, T.F. (1995). *Critical factors in the art of battle command.* (ARI Study Report 95-1). Alexandria VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A290 858)

Michel, R. (in preparation). *Measuring the structure of battlefield knowledge.* ARI Research Note.

Pounds, J. and Fallesen, J. J. (in preparation). *Problem solving of mid-career Army officers.* ARI Technical Report.

Solick, R.E., Spiegel, D.K., Lussier, J.W., & Keene, S.D. (1997). *Visualization and judgmental forecasting of simulated battles.* (ARI Research Note 97-08). Alexandria VA. US Army Research Institute for the Behavioral and Social Sciences. (AD A328 489)

Zacharias, G. L., Illgen, C., Asdigha, M., Yara, J., and Hudlicka, E. (1997).VIEW: *Visualization and Interactive Elicitation Workstation - A Tool for Representing the Commander's Mental Model of the Battlefield.* (ARI Research Note 97-01). Alexandria VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A322 458)

Dr. Lussier has worked as a psychologist at the Fort Leavenworth Research Unit of the Army Research Institute since 1984. Research interests include command and control, group planning and problem solving and battlefield thinking skills. He received an A.B. degree in Psychology from Columbia University and a Ph.D. in Experimental Psychology at Fordham University. Lussier served with the US Army as an interrogator with the XVIII Airborne Corps. He holds the title of expert with the United States Chess Federation.

AN ASSESSMENT OF THE FUNCTIONAL CAPABILITIES OF FOUR VIRTUAL INDIVIDUAL COMBATANT SIMULATORS

Robert J. Pleban
Infantry Forces Research Unit
Fort Benning, Georgia

Abstract

This research describes the results of an independent assessment of the functional capabilities of four prototype virtual individual combatant (VIC) simulators. Infantry soldiers were given the opportunity to operate each VIC in a series of squad-based scenarios requiring the performance of both individual and collective tasks in either a desert or urban setting. The results indicated that the more realistic the action or equipment used in the VIC, and the more reliable the VIC, the more the soldiers liked that system. The data collected from this research provide an important first step in the development of a set of dismounted infantry requirements for manned simulators that will support the integration of the individual soldier into the virtual battlefield.

Introduction

The Army has made a strong commitment to Distributed Interactive Simulation (DIS) as the major resource for conducting collective training and for planning and rehearsing military operations. The current DIS training system, Simulation Networking (SIMNET) and its successor, the Close Combat Tactical Trainer (CCTT), provide effective training for soldiers fighting from vehicles, but not for individual dismounted soldiers.

The introduction of the individual into the virtual battlefield had to be deferred because the number and complexity of models required to represent even a modest force of individual combatants exceeded the capacity of affordable real-time computing resources. However, rapid progress in the area of virtual environment systems has brought new technology to bear in addressing the complex issues of individual combat simulation. These systems immerse individual participants into synthesized surroundings through their own direct sensory experience. The resulting experience is one of personal presence and direct control of behaviors in the virtual world, rather than controlling tools or equipment which, acting as mediators, translate the individual's actions into observable effects in the virtual environment (Jacobs et al., 1994).

If this new technology is successful it provides for the possibility for expanding the capabilities of distributive interactive simulation to support mission proficiency training, mission planning, and mission rehearsal for the individual. This is important since the Army has assigned prime roles to airborne, airmobile, and light infantry divisions, and to Special Operations Forces (SOF), whose combatants fight primarily on foot (Jacobs et al., 1994).

In April 1997, the U.S. Army Research Institute (ARI) Infantry Forces Research Unit (IFRU) agreed to participate in a cooperative research project with the ARI Simulator Systems Research Unit and Lockheed Martin Corporation. The primary role of the IFRU in this project was to provide an independent assessment of the functional capabilities of four prototype virtual individual combatant (VIC) simulator technologies. The assessment took place during the Dismounted Warrior Network (DWN) User Exercises that were conducted over a three-week period at the Dismounted Battlespace Battle Lab (DBBL) at Fort Benning, Georgia. The User Exercises were only one part of the multi-phase DWN project. The primary objective of the project was the generation of a simulation task analysis document for Dismounted Infantry (DI) components. The focus of the User Exercises was to identify the strengths and weaknesses of the VICs from the perspective of the infantry soldier and to use these findings, in conjunction with engineering data collected earlier, to guide the development of future generation VICs and simulator systems requirements in general.

Method and Procedure

Eight soldiers from Fort Benning, Georgia participated in the virtual reality research. Soldiers received approximately one week of training in which they were briefed on the major characteristics and functional aspects of each VIC simulator. They were then given the opportunity to operate each VIC in a series of scenarios based on an infantry squad performing selected individual and collective tasks in either a desert or urban setting. The actual User Exercises lasted for two weeks. Paper-and-pencil questionnaires designed to assess the functional capabilities of the VICs across various dimensions were administered at selected times throughout the exercises. A structured interview was conducted with the soldiers at the end of the User Exercises.

Results

All VICs had certain strengths and weaknesses in their approaches. A key factor that emerged was the level of realism provided by the system. The more realistic the action (e.g., using a treadmill to actually walk versus using a joystick) or equipment (e.g., actual demilitarized weapon versus mockup), the more the soldiers liked the system. System reliability was another important theme that affected how well a particular VIC system was received by the soldiers.

Discussion

An important consideration in the development of future generation VICs is the specific purpose (s) to be served by these systems. For example, if the purpose of these VICs is mission rehearsal, how important is it that all actions be simulated in the VIC exactly the same way as the individual would perform them in the real world? Soldier ratings of these VICs were based, it seems, more from the perspective of systems for training individual soldier skills. In this instance, the realism factor makes sense. But is this really the best use of this technology for the infantry soldier? It may be that several VIC systems are needed; a system requiring realistic soldier actions and a more synthetic system for mission planning and rehearsal tasks. These issues will have to be addressed in the next set of User Exercises to ensure that the future generation VIC system(s) provide the maximum training value for the infantry soldier. The data collected from this research will provide an important first step in the development of a set of dismounted infantry requirements for manned simulators that will support the integration of the individual soldier into the virtual battlefield.

Dr. Pleban received his Ph.D. from the University of Georgia in social psychology with a co-major in industrial and organizational psychology. Besides working at the Army Research Institute-Infantry Forces Research Unit at Fort Benning, he has also worked as a research psychologist for the Walter Reed Army Institute of Research. Dr. Pleban has worked on a variety of research projects including basic research on the effects of varying work-rest cycles on cognitive performance and mood state in sustained operations; special forces assessment and selection; and the development and validation of effective multi-media distance learning strategies for infantry courses. His current research is on the effectiveness of virtual environments for training individual dismounted soldiers.

DEVELOPMENT OF A FAKING-RESISTANT TEMPERAMENT MEASURE: THE ASSESSMENT OF INDIVIDUAL MOTIVATION (AIM)[1]

Mark C. Young
Selection and Assignment Research Unit
Alexandria, VA

Abstract

The Army has had a growing need for new measures that can be used to augment education credential for predicting attrition risk and motivational components of performance. In the 1980s, the Army developed a new self-report temperament measure, the Assessment of Background and Life Experiences (ABLE). Although ABLE performs well in research contexts, it is highly susceptible to the effects of faking and coaching. For this reason, the operational use of ABLE in a large-scale implementation - such as for preenlistment screening - is not considered viable. The purpose of this research was to develop a prototype faking-resistant temperament/ biodata measure of ABLE constructs that could be used operationally for preenlistment screening. This new prototype measure is called the Assessment of Individual Motivation (AIM).

AIM was developed to measure six of the seven constructs measured by ABLE: Dependability, Adjustment, Athletic Interest (Physical Condition), Dominance, Achievement, and Agreeableness. AIM items were constructed using a forced-choice format with quasi-ipsative scoring to help place constraints on faking. A social desirability scale was also developed to detect respondents' efforts to inflate their scores by "faking good."

Research results indicate that AIM adequately measures ABLE constructs while being less susceptible to faking. As expected, AIM scores were found to predict training attrition, and the shape of this test score/attrition relationship was very similar to that which has been observed for ABLE. AIM shows promise as a preenlistment screening tool that could be used to reduce attrition and increase performance among selected applicants. Research plans needed to prepare for a possible AIM implementation are now being made.

The Army has had a growing need for new measures that can be used to augment education credential for predicting attrition risk and motivational ("will do") components of performance. One such measure, the Assessment of Background and Life

[1] The author would like to thank Len White for his contributions to this research. Dr. White is currently employed at the Office of Personnel Management in Washington, DC.

Experiences (ABLE), was developed out of the Army's Project A in the 1980s. ABLE is a self-report temperament measure that predicts both attrition and "will do" aspects of performance. "Will do" performance dimensions include personal discipline, effort and leadership, and physical fitness and military bearing.

Although ABLE performs well in research contexts, it is highly susceptible to the effects of faking and coaching. High levels of faking on ABLE have consistently been shown to attenuate its validity against attrition and job performance. For this reason, the operational use of ABLE in a large-scale implementation - such as for preenlistment screening - is not considered viable.

The purpose of this research was to develop a prototype faking-resistant temperament/biodata measure of ABLE constructs that could be used operationally for preenlistment screening. This new prototype measure is called the Assessment of Individual Motivation (AIM).

Approach

AIM was developed to measure six of the seven constructs measured by ABLE: Dependability, Adjustment, Athletic Interest (Physical Condition), Dominance, Achievement, and Agreeableness. AIM items were based heavily upon ABLE content. However, unlike ABLE, AIM items were constructed using a forced-choice item format with quasi-ipsative scoring to help place constraints on faking. In this scoring approach, an individual's scores on a given construct are partially (but not completely) dependent upon his or her scores on other constructs. This partial dependency prevents an individual from getting the highest possible scores on all constructs at the same time. A social desirability scale was also developed to detect respondents' efforts to inflate their scores by "faking good."

AIM was developed and refined over a four-year period using test data from approximately 5,000 Regular Army receptees. During this period there have been seven iterations of test refinement. Each iteration has consisted of (1) administering an AIM form to Army receptees, (2) performing item analyses, and (3) revising the AIM form in preparation for another test administration. During each iteration of test refinement, scale items which did not contribute to coefficient alpha reliability were revised. Also, items were revised (or dropped) when response alternatives had high endorsement rates (i.e., 70%+), or when content items had high correlations with the AIM social desirability scale. The goal here was to minimize the impact that social desirability would have on item responding.

During some iterations of test administration, ABLE was administered along with AIM. This was done so that ABLE constructs could be used as marker variables for assessing AIM's convergent and discriminate validity.

During other iterations of test administration, subsamples of respondents were given instructions to try to inflate their scores on AIM by faking good. In one sample we also coached examinees on how to score high on AIM. Data from these subjects were used in assessing items developed for measuring social desirability. These data were also used for evaluating AIM's resistance to faking.

Test respondents from selected administrations were tracked so that their attrition status during initial entry training could be established. These data were used to provide evidence of AIM's validity as an attrition predictor. Test respondents were also asked a series of questions pertaining to past use of illegal drugs, criminal behavior, and getting into trouble at work. The responses to these questions were used for creating a delinquency scale for use in validating AIM.

Findings

Research results indicate that AIM adequately measures ABLE constructs while being less susceptible to faking. The internal consistency reliability of AIM content scales (coefficient alpha ranges from .54 to .70) is someone lower than that of ABLE scales. This may be partially due to the quasi-ipsative scoring used with AIM. The AIM's adaptability composite, developed for predicting overall attrition risk, has good internal consistency reliability (coefficient alpha = .77).

As expected, AIM adaptability scores were found to predict early attrition, and the shape of this test score/attrition relationship was very similar to that which has been observed for ABLE. This finding has been consistent across independent samples of receptees. Correlations between AIM and other measures were also consistent with previous ABLE findings. The AIM adaptability composite had a low correlation with ASVAB. AIM Dependability, Adjustment, and Achievement all had the expected relationships with self-reports of conceptually related behaviors (i.e., drug use and criminal activities, effective coping skills, and getting into trouble at work).

Findings from one faking experiment showed that AIM's validity as an attrition predictor was maintained even when respondents were asked to inflate their scores on the test by faking good. In contrast, ABLE scores had no validity under this faking condition. Although these findings suggest that AIM is more faking resistant than ABLE, this experiment needs to be replicated on a larger sample using a more current version of AIM.

AIM scale scores were shown to be comparable for both male and females as well as whites and minorities. Based on these findings (and consistent with ABLE findings) there is no evidence that operational use of AIM as a screening tool would result in adverse impact.

Discussion

AIM shows promise as a preenlistment screening tool that could be used to reduce attrition and increase performance among selected applicants. In response to briefings provided by ARI, LTG Vollrath, Deputy Chief of Staff for Personnel, has expressed interest in using AIM as a screening tool for reducing enlisted attrition. However, before an implementation decision can be made, more research will be needed. In the next phase of AIM evaluation, larger and more representative samples of receptees will be required. Research plans needed to prepare for a possible AIM implementation are now being made.

Dr. Young completed an M.S. in experimental psychology from Villanova University in 1981. He then went on to complete a Ph.D. in organizational psychology from Georgia State University in 1987. His specialty there was in personnel psychology and measurement. In 1988, Dr. Young started working as a research psychologist with the U.S. Army Research Institute in Alexandria. Most of his tenure there has been in the Selection and Assignment Research Unit where he has investigated the use of temperament and biodata measures in adaptability screening.

THE EFFECTS OF THE AH-64A PILOT'S NIGHT VISION SYSTEM ON THE PERFORMANCE OF SEVEN SIMULATED MANEUVER TASKS

John E. Stewart II
Rotary-Wing Aviation Research Unit
Fort Rucker, AL

Abstract

The Pilot Night Vision System (PNVS) allows the pilot of the AH-64A Apache helicopter to fly and navigate effectively at night. The PNVS restricts field of view (FOV) and degrades visual acuity, two factors which should adversely affect pilot performance during night operations, especially for flight maneuvers that rely on peripheral cues. Participants, all AH-64A instructor pilots, flew a mission profile in a high-fidelity simulation of the AH-64A. The profile consisted of seven routine maneuvers, which were flown first under daytime (baseline) conditions, using a rear projection display with a 174o FOV, then under night conditions with a simulated PNVS display (40o FOV). Automated performance measures were the main dependent variables. It was expected that pilot performance degradation would be significant on hovering turns, takeoffs, and landings, and negligible on stationary hovering and hover taxiing. Significant performance degradations were found for all maneuver tasks. Applications of the findings to training and pilot workload standards are discussed.

Background

Thermal imaging night vision system. The AH-64A Apache helicopter is equipped with a night vision system (NVS), in a nose-mounted turret. It consists of two subsystems: the Pilot's Night Vision System (PNVS) and the Target Acquisition/ Designation System. This research focuses on the PNVS, a thermal imaging system with imagery and symbology presented via a helmet mounted display (HMD), the integrated helmet and display sight subsystem (IHADSS). The PNVS is slaved to the IHADSS so that it follows the motions of the pilot's head. It provides a 40o horizontal x 30o vertical field of view (FOV). The pilot's eyepoint is displaced to the PNVS by approximately 1.75 m (copilot-gunner's cockpit) and 3.50 m (pilot's cockpit). The IHADSS display is monocular, with the eyepiece mounted on the right side of the helmet, making it unique among NVS systems, most of which are binocular.

PNVS effects on performance. There is little empirical research on how the PNVS affects pilot performance (Foyle & Kaiser, 1991; D.C. Hart, 1994). S.G. Hart (1988) discusses PNVS characteristics that are likely to degrade performance. These include binocular rivalry, off-axis tracking, and eyepoint displacement, which may result in distorted visual motion cues. At night the pilot sees the instrument panel and the visual scene through the left eye. The right eye sees thermal imagery of the outside visual scene, from the nose of the aircraft. It seems likely that many of the problems with the PNVS are due to motion illusions and the lack of adequate visual motion cues . Because of the restricted FOV, we would expect performance degradation to be greatest for those tasks in which the pilot relies heavily on peripheral vision for velocity and motion cues.

Method

Participants and Design

The experiment was conducted at the U.S. Army Research Institute Simulator Training Research Advanced Testbed for Aviation (STRATA) at Fort Rucker, Alabama. The sample consisted of 11 rated AH-64A aviators. Nine had no previous exposure to STRATA. The copilot/ gunner cockpit was used in this experiment, because it was equipped with an IHADSS.

Procedure

Pilots performed seven maneuvers (hover, hover taxi, hovering turns, hovering out of ground effect (100 ft), normal takeoff, confined area landing and visual meteorological conditions (VMC) approach and landing to a hover) to aircrew training manual (ATM) standards. The mission profile was "flown" first under baseline (daylight), then under PNVS conditions. Independent variables were the two visual display conditions (baseline/ PNVS). Daylight operations were simulated using a rear-projection color display with a 174o FOV. For simulating night operations, the three rear-projection screens were dark and the pilot able to see the out-the-cockpit visual scene only through the IHADSS. Two retired Army helicopter pilots served as raters. A (5-point) Likert-type rating form was adapted from Stewart (1994). The following automated performance measures (APM)s were used:), airspeed (kt), drift (m), heading (degrees), lateral cyclic pitch control displacement (in), pitch and roll (degrees), rate of climb/descent (fpm, acceleration/ deceleration ((velocity), turn rate (degrees/sec). Data bearing upon pilots' experience in the AH-64A were collected, via questionnaire.

Results and Discussion

Performance Based on Real-Time Ratings

A Pearson r of .73 ($p < .005$) indicated acceptable inter-rater reliability. As expected, overall performance was better in the baseline condition ($F(1,9) = 19.14$, $p < .002$). All maneuvers were performed to ATM standards. Deficits due to PNVS flight were expected to be greatest for those maneuvers involving dependence on peripheral cues (hovering turns, normal takeoff, VMC approach and landing, confined area landing). Differences for the first three previously-mentioned maneuvers approached or exceeded $p < .05$. No other differences were significant.

Performance based on Automated Performance Measures

APM data were examined to investigate aircraft state differences in baseline vs. PNVS flight. APMs revealed deficits across all maneuver tasks during PNVS conditions with the predominant difference being greater excursion about the roll axis. The induced roll could have been due to differences in FOV, asymmetric aircraft attitude while hovering, and the "instinctive" need on the part of pilots to see a level horizon. The AH-64A hovers rolled approximately 3o to the left. The tilted scene, as viewed through the PNVS, will remain tilted, regardless of pilot head orientation. The pilot can attempt to level the scene by rolling the aircraft to the right. This causes drift, necessitating a roll back to the left. Thus begins the cycle of induced roll oscillation.

It appears that for all maneuver tasks in the mission profile, a tradeoff for enhanced night vision capability was diminished aircraft control, which is especially noteworthy when one recalls that all participants were experienced IPs, who operate routinely in PNVS mode. This was possibly due to the restricted FOV in the PNVS condition. The visual scene through the PNVS, however, differed from the baseline, day scene on several other dimensions besides FOV. The complexity of the system makes this post hoc explanation somewhat simplistic

Roll during takeoff may have been induced by relative absence of peripheral (relative motion) cues, making it more difficult for the pilot to maintain a constant alignment with a reference point on the ground. As was seen in the course of the normal takeoff, control of this augmented roll became critical to successful performance.

Data bearing on pilot experience (AH-64A pilot hr, NVS flight hr, elapsed time since last NVS flight, and total hours in the AH-64A Combat Mission Simulator) were correlated with APMs. Combat Mission Simulator (CMS) hr seemed to be the best predictor of performance on the APMs, correlating significantly and in the expected direction with at least one APM for six out of the seven maneuver tasks. The small sample size may have limited the number of significant correlations with this experience variable. CMS hr correlated significantly with real-time performance ratings in the baseline condition for confined area landing and VMC approach. CMS is an operational training simulator, technically different in many ways from STRATA. These significant correlations could indicate convergent validation of two simulation approaches to the same aircraft.

Applications

This experiment provides objective data which indicate performance degradation of AH-64A handling characteristics for PNVS flight. These findings have practical implications for training, flight safety and workload measurement. Army Regulation 95-3 provides a rule of thumb for crew endurance under day and night vision (NV) device conditions. Under the current rubric, one hour of NV flight is considered equivalent to 2.3 hours of flight under day conditions (Department of the Army, 1990). This ratio is for NV devices in general. It does not distinguish PNVS from Night Vision Goggles (NVG)s. An experimental approach, using simulations of different NV systems, could yield a more precise, empirically-based foundation for determining flight time equivalencies. It could also pinpoint specific maneuvers in which NV device use poses the greatest challenge to safety and mission performance.

References

Department of the Army (1990). *Aviation: General provisions, training, standardization, and resource management* (Army Regulation 95-3).

Foyle, D.C. & Kaiser, M.K. (1991). *Pilot distance estimation with unaided vision, night vision goggles, and infrared images.* SID International Symposium Digest of Technical Papers, XXI (pp. 314-317). *Los Angeles: Society for Information Display.*

Hart, D.C. (1994). ADS-33C flight maneuvers validation in a degraded visual environment. *American Helicopter Society 50th Annual Forum Proceedings* (pp. 1141-1154). Alexandria, VA: American Helicopter Society.

Hart. S.G. (1988). Helicopter human factors. In E.L. Weiner & D.C. Nagel (Eds.), *Human factors in aviation* (pp. 591-638). San Diego: Academic.

Stewart, J.E. (1994). *Using the backward transfer paradigm to validate the AH-64 Simulator Training Research Advanced Testbed For Aviation.* (Res. Rep. No. 1666). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. AD A285 758.

Dr. Stewart received his Ph.D. in Social Psychology from the University of Georgia at Athens in 1973. He served as Assistant Professor at Mercyhurst College from 1973-79. From 1979-84 he developed methodologies for evaluating the effectiveness of various federally-funded social service/transportation programs for the State of Arizona. From 1984-85 he was a Research Psychologist with the U.S. Air Force Human Resources Laboratory at Williams Air Force Base, where he conducted research on the training effectiveness of an air refueling part-task trainer for C-5 and C-141 aircrews. He transferred to the Army Research Institute in 1985, where he was assigned to the Systems Research Laboratory. During that time he pilot-tested methodologies for forecasting maintenance manpower requirements for emerging Army weapons systems. In 1989 he transferred to the Rotary Wing Aviation Research Unit at Fort Rucker. His current research interests are in the areas of simulator validation, transfer of training, training system development and evaluation, and the development of metrics for performance measurement.

COMMAND ENTITIES: COGNITIVE BEHAVIORS FOR COMPUTER GENERATED FORCES

Dr. Philip Gillis
Simulator Systems Research Unit
Orlando, FL

Abstract

The Command Entities Cognitive Behaviors for Computer Generated Forces project is an effort to produce better methods to represent human performance variability in battlefield simulations. Artificial command entities have been developed that are subject to the effects of stress, fatigue, situational influences, training, experience, individual differences and other factors. An attempt is then made to validate the behaviors of these command entities by means of placing them in command of a Blue Force (BlueFor) Task Force that faces an opposing force, an OpFor, with both BlueFor and OpFor battlefield entities input from actual National Training Center data. The major research objectives of this project include investigations of: 1) the situational awareness of battlefield actions and events by both humans and computers and how that process can be modeled, 2) the nature of "realism" in CGF behavior, and 3) human performance variability on the battlefield.

Problem

The basic problem addressed by this research concerns the need for more intelligent and realistic behavior by Computer Generated Forces (CGF) Command Entities (CE's) in Distributed Interactive Simulation (DIS) through the utilization of better human performance cognitive modeling R&D. Realistic cognitive modeling class specifications for CGF have yet to be developed, especially for large scale simulations, such as WARSIM 2000 and JSIMS.

This lack of research impacts the accurate behavior of CGF entities, and it also impacts the human resources required for CGF scenario development and exercise control in DIS. They are inappropriately demanding and may be offloaded by more intelligent CGF behaviors.

In addition, the development of human performance cognitive models (HPCM) for CGF entities needs to be validated in a realistic simulation setting. ARI has access to an extant DIS testbed, C3SIM, and also has access to numerous NTC/CTC training missions that provide a means for the empirical validation of the HPCM.

Description

The overall objectives of this effort are to identify, develop and evaluate: 1) a minimalistic, unified, expanded theory of human behavior on the battlefield; 2) the nature of "realism" in CGF behavior; and 3) a human performance cognitive model based upon the effects of training, experience, combat stress, fatigue, confidence-building events, morale, individual differences, situational influences, miscommunication, and the physiological and psychological effects of direct/indirect fire suppression, and heat and ballistic injury for Computer Generated Forces Command Entities.

The HPCM is being evaluated through the hosting of the cognitive model in an ARI DIS simulation, C3SIM. Direct access to numerous NTC/CTC training exercises are being used to test the validity of the HPCM.

Background

In Phase I, a Battalion Command Entity (BCE) has been developed by ARI for the command and control of Bluefor units, matched against Opfor units, all taken from actual Task Force training data from the National Training Center. The resulting simulation, C3SIM, utilizes maps, terrain data, missions, BlueFor and Opfor positions, strengths, and movements directly derived from NTC data. All CE's utilize METT-T and the command decision-making process for the cognitive modeling of C3 operations. In addition, the BCE cognitive model, in omniscient mode, uses AI-technology for purposes of situational awareness enhancement and decision aiding. These algorithms include techniques from expert systems, fuzzy logic, knowledge-based systems, and semantic nets research. The cognitive modeling architecture is completely reconfigurable and can easily be modified to accommodate the changes in the BCE's behaviors due to the aforementioned stress and readiness factors.

Major Accomplishments of Phase I:

Phase I of this effort resulted in two major accomplishments. First was the development of a human performance cognitive model, HPCM Ver. 1.0. This software prototype primarily utilized Walter Reed Army Institute of Research data on the effects of continuous operations and stress on individual performance. At the heart of the model is a cognitive reservoir that maintains a balance of effective performance units. An effectiveness ratio was produced as the product of interactions between battlefield stressors, performance circadian rhythm, performance use, confidence building events, and training and experience variables. This effectiveness ratio was then divided by the number of potential courses of action that a simulated command entity could decide between, producing a probability of a correct decision. A timing variable was also introduced; it utilized the HPCM "effectiveness" ratio in order to produce a time lag for the selection of a course of action by a possibly "fatigued and stressed" BCE.

The second major accomplishment of this project was the refinement of a PC-based, National Training Center mission replay simulation that can be used to test the effectiveness of the HPCM outputs. The resultant simulation, C3SIM, is a reconfigurable, constructive simulation that utilizes Task Force strength Bluefor and Opfor Battlefield Operating Systems that may be directly input from archived NTC data. C3SIM features mission replay or mission modification modes for both Bluefor and Opfor Battalion Command Entity control. Thus, C3SIM may be used to obtain baseline data related to human performance variables or it may be used to observe the effects of sleep deprivation, battlefield stressors, experience and training, and other factors.

Phase I/Preliminary Findings:

1. The following factors **degrade** performance, as is manifested in the BCE's selection of a course or action and the BCE's timeliness of decisions made.

   a. Sleep deprivation — A factor such as sleep deprivation is a quantifiable variable; its effects on the performance reservoir is in the form of a continuous variable.

   b. Performance use — This is also a real number that varies over time. As time goes by, the BCE's performance use results in the decrement of the effectiveness variable, which consequently decreases the BCE's probability of making a correct decision.

   c. Situational Influence

      (1) Battlefield Stress Events

      (2) Time Pressure

   d. Insufficient Experience

These effects modulate the effectiveness variable. Such effects take into account the frequency and severity of stress events, and time pressure.

2. The following factors may **enhance** performance in quantifiable units—they serve to modulate, in a positive manner, the effects of stress:

   a. Good—Superior Experience

   b. Confidence Building Events

In phase I, the experience variable and confidence building effect served to modulate the effectiveness variable either by means of decreasing the stress effect.

3. Several factors affect performance in non-quantifiable, though significant manners; for example, a BCE may choose a riskier COA that may produce a better outcome than a more risk aversive choice:

a. Aggression versus risk-aversive

b. Emotion/motivation

Additional issues that need to be examined here are factors such as, when a BCE becomes fatigued, does it take more chances?

4. There are conflicting factors that need more examination. For example, as stress increases, intelligence becomes less important and experience becomes more important.(Locklear, Powell, and Fiedler, 1988)

References

Locklear, Powell, & Fiedler (1988). The effects of individual experience and intelligence on decision performance, University of Washington report.

Rosenbloom, P.S., Laird, J.E., Newell, A., (Eds.) 1993. The Soar Papers: Research on Integrated Intelligence. Cambridge, MA:MIT Press

Dr. Gillis is a Cognitive Research Psychologist, working for the Army Research Institute and STRICOM in Orlando. He has published numerous articles/conference proceedings on the issues of AI, training, and simulation. Dr. Gillis is presently conducting research in the area of realism in computer-generated forces and human performance on the battlefield, under the auspices of DARPA, the National Simulation Center, and the U.S. Army Artificial Intelligence Center at the Pentagon.

PREMO: ACCELERATING MOBILIZED SOLDIERS' REACQUISITION OF SKILLS

Mark A. Sabol
Advanced Training Methods Research Unit
Alexandria, VA

Abstract

Videotaped demonstrations of three common soldiering tasks were shown to 100 members of the Individual Ready Reserve. Days later, they performed tasks they saw significantly better than matched tasks. For soldiers with two years or more of active duty, higher AFQT scores correlated with better performance on tasks not demonstrated; exposure to the taped demonstrations eliminated this AFQT advantage. No disadvantage was found for those out of active duty service longer than others. A skill reacquisition curve is described; time spent on refresher training, expressed as a portion of original training time, appears to predict refresher training success.

Introduction

Studies conducted by ARI researchers in the 1980s (e.g., Hagman & Rose, 1983; Rose, Hagman, Radtke, & Shettel, 1985) determined what characteristics make a task, when learned by the average soldier, susceptible to a rapid decline in quality during periods of non-practice. These task variables are now well known, and we can confidently predict rates of skill decay in average task performance. There is, in fact, an Army manual, User's Manual for Predicting Military Task Retention, that describes in detail how to conduct the interviews with subject matter experts which yield the data needed to make such predictions. Our Acquisition and Retention of Cognitive Skills team has just concluded a research program designed to complement that earlier one. The new program asked what characteristics make an individual soldier, when performing a typical military task, susceptible to rapid forgetting during periods of non-practice. We sought first to identify the demographic and experience variables that would predict the decay rate for individuals performing the average procedural task. We later sought to predict the rate at which different individuals can reacquire skills as they experience what the Army calls "rapid train-up."

The Individual Ready Reserve (IRR)

Soldiers in the IRR are ideal subjects for such research. These are people who leave active duty but have time left on their contractual obligation to the Army and are thus liable to being called back to active duty in an emergency. Unlike people in the Army Reserve or National Guard, who are assigned to units and get periodic training opportunities with those units over weekends and summer exercises, members of the IRR are, by definition, not assigned to units and receive no such training opportunities. They are, therefore, asked to retain skills over months and even years without practicing those skills in the setting in which they were learned.

Beginning before Operation Desert Storm, we have observed such soldiers on several occasions when they were called to active duty, usually as part of mobilization exercise. Our performance measures have been written job knowledge tests, hands-on task performance, or both. Separate studies of skill retention have been conducted on soldiers whose original training was in a variety of military specialties: radio operators (Sabol, Chapell, & Meiers, 1990), quartermasters (Wisher, Sabol, Sukenik, & Kern, 1991), combat engineers (Kern, Wisher, Sabol, & Farr, 1993), vehicle mechanics (Sabol, Kern, Eidelkind, & DiMarino, 1993) field medics (Wisher, Sabol, Maisano, Knott, Curnow, & Ellis, 1996), masonry/carpentry specialists, air defense missile crews, and military police (Wisher, Sabol, & Ozkaptan, 1996).

The demographic/experiential variables we used in these studies include: 1) the soldier's general aptitude for learning (score on the Armed Forces Qualification Test, AFQT), 2) the soldier's previous level of knowledge acquisition (score on the Army's old Skill Qualification Test, SQT), 3) extent of prior practice at the tasks (a soldier's time on active duty in the relevant military specialty), 4) retention interval (a soldier's time since leaving active duty), and 5) partial practice during the retention interval (the self-rated similarity of the soldier's civilian job to the relevant military specialty). These characteristics of individual soldiers were correlated with measures of written job knowledge and hands-on procedural skills both before and after the rapid train-up provided during mobilization exercises.

In each study, we found that the best predictor of resistance to knowledge decay (on tests taken before the rapid train-up) was "overlearning," a combination of aptitude (AFQT score) and strength of original learning (SQT score or the soldier's active duty time). That is, the soldiers who retained knowledge best were those who had an opportunity to use that knowledge beyond the original learning experience. This overlearning predictor was followed closely in strength by the relevance of the soldier's civilian occupation. Performance after reacquisition was best predicted by the aptitude (AFQT) measure. We have consistently found only a small detrimental effect due to increase in retention interval. That is, soldiers who have been away from active duty for long periods do show a loss of knowledge and skill, but not to the extent most people assume; the other factors were found to be far more powerful.

The CALL FORWARD (CF) Exercises

Our interest has focused lately upon hands-on performance. In the mobilization exercise known as CF95, we followed mobilized IRR soldiers who had been field medics or air defense crewmembers in the Army as they experienced rapid train-up in both common soldiering tasks and the specialized tasks of their old MOS. From data collected in this exercise, we developed a preliminary version of a "skill reacquisition curve," which relates the percentage of tasks on which soldiers are expected to receive a "Go" not to characteristics of the tasks nor of the soldiers themselves, but rather to an aspect of the training procedure: the ratio of time spent on the tasks during the rapid train-up to the time originally spent training those same tasks (as specified in the Army's program of instruction for the appropriate MOS).

In CF97, we followed up on an effect revealed in CF95: the influence of soldiers' civilian jobs upon their retention of skills. Soldiers mobilized from civilian jobs in which they perform tasks similar to their military tasks (for example, field medics who held civilian jobs as hospital worker) show little decay in their ability to perform those military tasks. Because the tasks involved are not identical, we think this effect is due not so much to continued practice as to the maintenance of an appropriate context. A civilian medical worker does not need time to reinstate the "frame of mind" necessary for his work as a field medic. We think that the typical IRR soldier can be helped to maintain or reestablish the correct context and frame of mind through modern technology. In particular, we have tested the idea that exposure to abbreviated training, presented at the right time, can remind the soldiers of their tasks in a way which triggers additional recall. The goal we have in mind is interactive, computer-based training which could be delivered to all IRR soldiers in the time between call-up notice and muster.

CF95 — SIGNIFICANT PREDICTORS OF MOS HANDS-ON PERFORMANCE

- Those in medical jobs in civilian life (N=33)
  - Diagnostic (pre-rapid-train-up): none
  - Post-train-up: none
- Those <u>not</u> in medical jobs in civilian life (N=79)
  - Diagnostic:
    - Aptitude (AFQT), $p < .001$
    - Length of Active Duty, $p < .001$
    - Medical Aspect of Civilian Job, $p < .025$
    - Rank, $p < .05$
  - Post-train-up:
    - Aptitude, $p < .025$

NOTE: Time out of service had no predictive value.

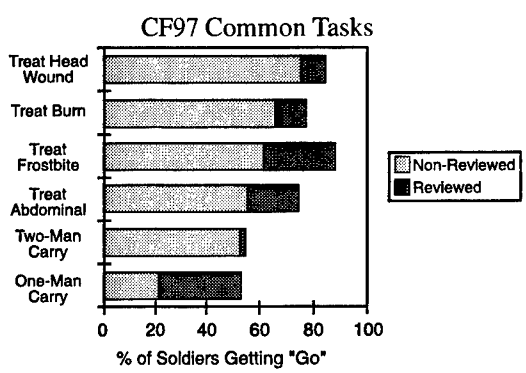
Method

In our procedure for CF97, this training was represented by a videotape review of four common tasks, three medical and one on map reading. We prepared two tapes showing two sets of such tasks being performed in accordance with the 1994 Soldier's Manual of Common Tasks, Skill Level I. We showed one video twice to half of the sample of 100 soldiers and the other video twice to the other half. All 100 were asked, several days later, to perform all eight tasks.

Several things went wrong along the way, the most important being changes made in the tasks by the drill sergeants doing the actual testing. These changes had the effect of increasing the likelihood that a soldier would perform the map interpretation task on one videotape correctly if he had been exposed to the opposite videotape. As this would obviously work against our hypothesis, we were forced to eliminate both map tasks from our task sample. Another problem involved the motivation of the IRR soldiers. Our initial questionnaire asked the soldiers for a self report of level of interest in relearning both common tasks and their respective MOS tasks. Ten soldiers reported very low motivation to relearn common tasks, and many of these soldiers performed poorly on those tasks. All ten were eventually dropped from the data analysis.

Results

That analysis revealed the same effects of personnel variables which we have found in earlier studies: 1) a highly significant improvement in performance on tasks shown on the videotape over tasks not shown; 2) an advantage for soldiers who had completed a full tour of active duty in the Army before entering the IRR, as opposed to those who entered after only a few months of training in their MOS; 2) a performance advantage for those whose aptitude (as measured be AFQT score) was above average, especially on tasks not reviewed on the videotape; and 4) no effect on performance of the length of time an IRR soldier had been out of the Army. In addition, we found that the data points for the "skill reacquisition function" obtained in CF97 fell precisely in line with the points from CF95.

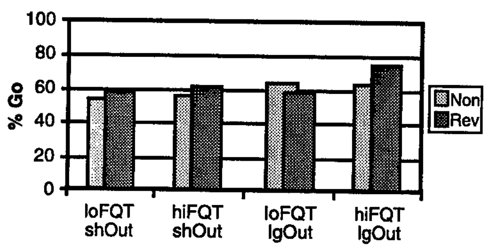


- Average 14% of original training time
- Average non-reviewed Go rate. 54%
- Average reviewed Go rate. 71%

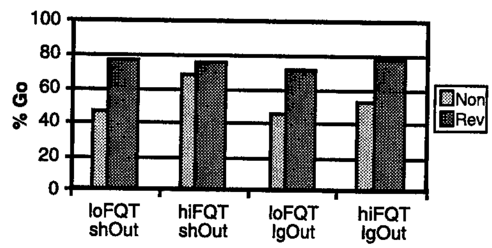
**p<.01
* p<.05

In summary, we found brief exposure to simple demonstrations of tasks sufficient to cause clear improvement in hands-on performance. As the time needed to present these demonstrations twice averaged about 5 minutes per task, this represents a dramatic savings (~85%) over the time spent on initial training. Yet nearly a third of the soldiers reported on a post-training questionnaire that the videotapes were "very useful," in that they "brought back a lot" the soldiers once knew about their Army tasks. These results encourage efforts to employ distance learning technologies to shorten rapid train-up time in future mobilizations.

Partial Tour (n=26)



Full Tour (n=64)

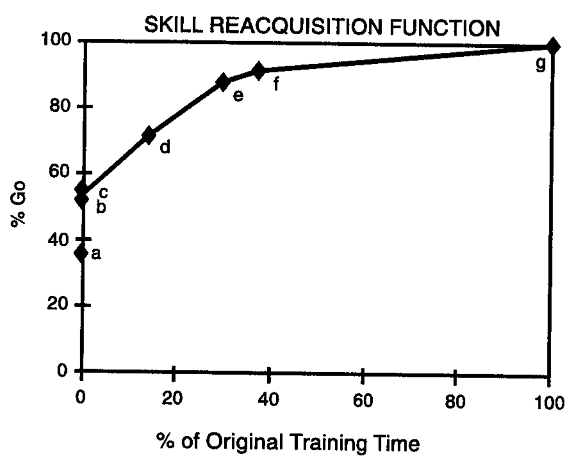


- Significant Effects:
  - Partial Tour: none
  - Full Tour: review ($p<.001$), AFQT ($p<.05$),
    - AFQT x review (marginal, $p<.15$)

Note: Time out of service again not important

SKILL REACQUISITION DURING RAPID TRAIN-UP

- Consolidation of data from CF95 and CF97; total n = 286



DATA POINTS

a	MOS Diagnostic testing, CF95 (2,900 observations)
b,c	Diagnostic CTT, CF95 (2,900 observations) & CF97 (250 observations)
d	Reviewed CTT, CF97 (250 observations)
e	CTT Post-Rapid Train-Up, CF95 (2,600 observations)
f	MOS Post-Rapid-Train-Up, CF95 (2,600 observations)
g	Full Instruction (assumed)

References

Hagman, J.D., & Rose, A.M. (1983) Retention of military tasks: A review. Human Factors, 25, 199-213.

Kern, R.P., Wisher, R.A., Sabol, M.A., & Farr, B.J. (1993) Reacquisition of skill by combat engineers mobilized from the Individual Ready Reserve, ARI Research Report 1667.

Rose, A.M., Hagman, J.D., Radtke, P.H., & Shettel, H.H. (1985). User's Manual for Predicting Military Task Retention, ARI Research Product 85-13.

Sabol, M.A., Chapell, L.G., & Meiers, C. (1990). Predicted Decay of Mobile Subscriber Equipment (MSE) Operator Skills, ARI Research Product 90-11.

Sabol, M.A., Kern, R.P., Eidelkind, M.A., & DiMarino, C.M. (1993). Effects of rapid training on soldier's occupational knowledge. Proceedings of the Military Testing Association.

Wisher, R.A., Sabol, M.A., Maisano, R.E., Knott, B.A., Curnow, C.K., & Ellis, J. (1996). Retraining medics for mobilization. Paper presented to American Educational Research Assn.

Wisher, R.A., Sabol, M.A., & Ozkaptan, H.R. (1996). Retention of "peace support operations" tasks during Bosnia deployment: A basis for refresher training. ARI Special Report.

Wisher, R.A., Sabol, M.A., Sukenik, H. & Kern, R.P. (1991). Individual Ready Reserve (IRR) call-up: Skill decay, ARI Research Report 1595.

Dr. Sabol is an old Indians fan from Cleveland. He earned a Bachelor of Science degree in Geology and a letter in fencing from the University of Notre Dame in 1968. His doctorate from Bowling Green State University (1979) is in experimental cognitive psychology, with a concentration in human memory. After teaching a half-dozen years in Pennsylvania, Nebraska (Creighton University), and Virginia, he came to ARI-Alexandria in 1985. His time away from work is spent mostly on his chores as president of the Slovak-American Society of Washington.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) February 1998	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) November 1997

4. TITLE AND SUBTITLE 1997 IN-HOUSE RESEARCHER COLLOQUIUM	5a. CONTRACT OR GRANT NUMBER
	5b. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Michelle M. Zazanis, Bruce Sterling, James W. Lussier, Robert J. Pleban, Mark C. Young, John E. Stewart II, Philip Gillis, Mark A. Sabol	5c. PROJECT NUMBER
	5d. TASK NUMBER
	5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: PERI-ZB 5001 Eisenhower Avenue Alexandria, VA 22333-5600	8. PERFORMING ORGANIZATION REPORT NUMBER
18. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue Alexandria, VA 22333-5600	18. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Special Report 35

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT *(Maximum 200 words)*

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) held a formal In-House Researcher Colloquium on 20 November 1997 in Alexandria, Virginia. The main purpose of the colloquium was to provide an opportunity for cross-unit discussion among ARI's more junior researchers. The eight researchers who presented research findings at the colloquium represented ARI's Armored Forces Research Unit, the Automated Training Methods Research Unit, the Fort Leavenworth Research Unit, the Infantry Forces Research Unit, the Organization and Personnel Resources Research Unit, the Rotary-Wing Aviation Research Unit, the Selection and Assignment Research Unit, and the Simulator Systems Research Unit. Each research topic was specifically selected by the Research Unit Chief as an example of the best of research being performed at the unit. This report provides brief summaries of the research and biographies of the researchers. It also serves as an example of the range of behavioral and social science research being addressed by in-house researchers at ARI as well as of the backgrounds of ARI's research staff.

15. SUBJECT TERMS

Peer Evaluations	Special Forces	Battle Staff Training	Thinking Skills
Battle Command Skills	Virtual Individual Combatant Simulators	Faking	Assessment of Individual Motivation
Temperament Measures	Night Vision	Computer Generated Forces	Skill Reacquisition

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT	18. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	36	Dr. Judith Brooks, 617-0345