EFFECTIVENESS EVALUATION FOR AIR COMBAT TRAINING

JAMES McGUINNESS, Ph.D.
JOHN H. BOUWMAN
Person-System Integration

and

JOSEPH A. PUIG
Naval Training Equipment Center

ABSTRACT

This paper addresses methods developed for evaluating factors involved in Air Combat Maneuvering (ACM) training. In the course of selecting and applying evaluation techniques, a unique situation for a transfer-of-training study was presented: a newly installed ACM simulator co-located with an ACM range. A common, objective performance measurement system was developed for the Air Combat Maneuvering Simulator (ACMS), designated Device 2E6, and the Tactical Aircrew Combat Training System (TACTS) range. The TACTS range was planned as the setting for studying transfer from the simulator to an operational situation.

INTRODUCTION

An early attempt to evaluate pilot trainees in air-to-air combat was known as "pin ball" (USAF, 1945). The "pin ball system" registered hits of frangible bullets on a manned, armor-plated, target aircraft. The system also provided a visual signal to the attacking pilot by turning on a strobe light in the nose of the target aircraft. The fidelity, realism and immediacy of objective feedback were high. As a gunnery trainer "pin ball" was a relatively effective training device, even though it probably did little for the pulse rate of the target aircrewman.

In 1967, NTEC began development of a gunnery practice system based on the use of inexpensive, eye-safe laser transmitters and receivers to simulate firing live rounds. Laser simulation has the advantage of providing an unlimited source of hazard-free "ammunition" at a negligible cost per round. Previous applications of this technology include the Laser Marksmanship Rifle Trainer and the Helicopter Door Gunnery Trainer.

The advent of air-to-air missiles in air combat maneuvering made the application of PIN BALL training techniques very inefficient, and required a technological sophistication beyond laser designators.

Training Air Combat Maneuvering (ACM) skills requires the exercise of a complex pattern of perceptual, psychomotor, physiological cues and procedural elements within a demanding tactical environment. You cannot train such a pattern in a simulator exclusively. Neither can you efficiently learn such a pattern in an aircraft exclusively. Training devices must be designed to complement actual training in aircraft. The increased complexity of fighter aircraft include multiple weapons for ACM employment, multiple systems for acquisition, two-man crews, additional functions required to operate weapons systems, and increases in aircraft performance which require quicker aircrew reaction time. All of these factors mandate increases in both flight time and simulator usage, and effective integration of training devices and aircraft into the fleet ACM training program structure. Improved training can have a force-multiplier effect in ACM, but we must bite the bullet. More training sorties in the air, undoubtedly, are required; and more simulator hours also will contribute to the training process. Maximum contribution of a simulator is contingent upon the integration of proper maintenance and instructor/operator support.

A Training Effectiveness Evaluation (TEE) was initiated by the Human Factors laboratory at the Naval Training Equipment Center (NAVTRAEQUIPCEN) to aid in defining, through objective measurement, how much and what type of integration is required. The TEE effort has been designed to provide data to help determine how well a training system produces a desired result. Such data must be forthcoming if we are to design for (1) optimal simulator/aircraft mix, (2) instructional strategies such as optimal sequencing, or (3) other significant factors that affect the amount and type of training required. The successful quantification of training effectiveness also will provide a data base which can be used for the specification and development of new or modified simulator systems.

The TEE took advantage of a unique situation for the development of such a data base. A recently installed ACM simulator co-located with an ACM range at NAS Oceana, Virginia, was reviewed for the feasibility of conducting a transfer-of-training study. A review of current ACM aircrew training programs, as well as a review of simulator and training device applications, suggested the need for a common Performance Measurement System (PMS) between the Air Combat Maneuvering Simulator (ACMS), designated Device 2E6, and the Tactical Aircrew Combat Training System (TACTS) range.

Device 2E6 is a high technology simulator consisting of two 40-foot diameter domes with an adversary aircraft image projected on the surface of the dome. The adversary image moves in response to inputs from an aircrewman in the opposing dome or in response to computer-controlled or console operator-controlled inputs. Simulated missile firing, as well as gun firing, are possible in the simulator and are accompanied by visual cues. This type of interaction can occur on a 1 versus 1 basis (1v1), on a 2 aircraft versus 1 aircraft basis (2v1), or on a 3 aircraft against one another basis (1v1v1).

The TACTS range permits objective measurements of aircraft spatial relationships, missile maneuvering envelopes and simulated missile firing in an environment that is as close to combat as is possible. The TACTS range instrumentation records selected movements and actions in specially instrumented aircraft as they interact in a "dogfight" over the Atlantic Ocean. The system tracks aircraft operating within an assigned airspace and computes and stores aircraft position, attitude, and weapons-related parameters in real time, permitting a three-dimensional depiction of the engagement to be monitored at a ground station. Radio communications are provided between the ground station supervising the exercise and the participating aircraft. Potentially hazardous flight conditions are automatically detected and brought to the attention of a training supervisor for appropriate action.

It was clear that if a common performance measurement system could be developed, the potential for a classical transfer-of-training (TOT) design and associated statistical analyses for the TEE would be enhanced. Thus, for the first time it appeared feasible to measure transfer of performance from simulator to ACM flight conditions as represented on the TACTS. The ACMS is a one-of-a-kind device and joint operation with TACTS yields quantitative air combat maneuvering performance data which are a direct function of pilot training and performance.

The first step in the TEE was to examine the TACTS range. TACTS was developed to provide aircrews with an objective means for improving missile envelope recognition. The TACTS provides training in a realistic but controlled environment. The trainee receives limited real-time airborne feedback and, later, more thorough "debrief" feedback concerning the effectiveness of weapons firing which are simulated by the TACTS computers at the ground station.

The TEE program also was timely in that it was able to directly support Commander Fighter Wing One (COMFITWING ONE) personnel who were in the process of developing a PMS for the TACTS range. The measurement system being developed for TACTS was entitled The Readiness Index Factor (RIF). It was based upon the Readiness Estimation System (RES) first developed by the Center for Naval Analyses. The RIF provided a measure of the spatial relationship of two or more interacting aircraft and considered the type of weapons employed.

DATA COLLECTION AND ANALYSIS

TACTS data were collected and analyzed using the RIF during two Fleet Fighter ACM Readiness Programs (FFARPs), which comprise a syllabus of "refresher" exercises (1v1, 2v1, 2v2, etc.) for fleet squadrons that span roughly three weeks. One FFARP involved airborne combat exercises employing friendly F-4 aircraft engaged with F-5 and A-4 adversary types. The other FFARP involved F-14 against F-5 and A-4 adversaries. The data indicated that the RIF was sensitive not only to differences in individual aircrews, but also to aircraft variables as well.

Figure 1 contains one comparison of such differences.

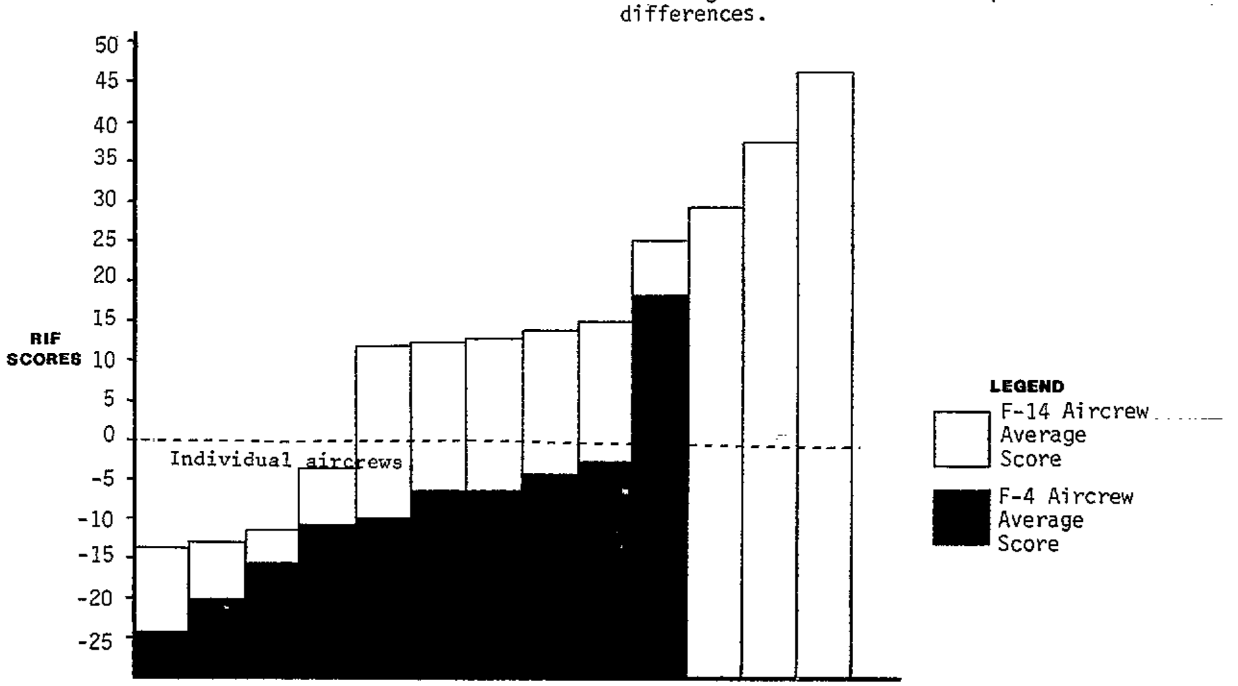


Figure 1. TACTICAL AIRCREW COMBAT TRAINING SYSTEM (TACTS) DATA

A RIF score above zero indicates that an aircraft is relatively more offensive than defensive in relationship to an adversary aircraft. A negative RIF score indicates the friendly aircraft is in a defensive posture. The RIF scores in Figure 1 indicate that, of the ten F-4 aircrews, only one was able to achieve an overall positive offensive average. On the other hand, all but four of the thirteen F-14 aircrews attained an overall positive offensive average.

Aircrews also were ranked according to objective RIF scores and these scores were compared to rankings obtained from adversary squadron pilots, ACM instructor grade sheets, and data from the "Blue Baron" data collection effort conducted on the TACTS range. The latter effort consisted of a combination of objective and subjective data, including parameters such as radar range, system lock-on parameters, visual sighting and kill/miss assessments. In-depth analysis of the type of shots taken revealed that a high percentage of missile shots fired were in the forward hemisphere. Further data analysis revealed that the RIF was not sensitive to forward hemisphere (i.e., all-aspect) missile firings.

ALL-ASPECT MANEUVERING INDEX (AAMI)

As a result of the first data collection, it was concluded that an all-aspect capability had to be incorporated in the RIF. Using the successful elements of the RIF, an All-Aspect Maneuvering Index (AAMI) measurement system was designed and developed for the next data collection series. This series involved data collected from training being conducted on the ACMS, involving participation of a number of F-14 and F-4 squadrons operating out of NAS Oceana.

The AAMI is a measurement system based upon a formula that incorporates range, antenna-train-angle, and angle-off-the-tail as the major variables which define the spatial relationship among interacting aircraft. The formula is weighted by a weapon range modifier which accounts for the influence of individual weapon systems such as the AIM-9L, AIM-7F, and AIM-9G. As with its RIF predecessor, an AAMI score above zero indicates that an aircraft is more offensive than defensive in relationship to an adversary aircraft. A high positive score is awarded when the aircraft achieves an optimal position with respect to the weapons load it has on board. In addition to providing graphical feedback, the AAMI describes a range of numeric indicators such as time to fire, time to first kill, missile type information, number of gun rounds, etc.

Figure 2 contains AAMI summary data from the simulator. All simulator engagements covered in this paper pitted an aircrew experienced in ACM against a computer-driven adversary. The data in Figure 2 reflect differences due to aircraft variables quite similar to the differences shown by previous RIF data collected on the TACTS range. For example, F-14 aircraft consistently outperform F-4 aircraft against adversary aircraft in a simulated environment. The individual differences and performance variations among the three F-4 aircrews and the two F-14 aircrews have been assessed in light of the historical records supplied by each aircrew. One point stands clear from an examination of the data in Figure 2: Post-test scores are much higher than Pre-test scores. This indicates that the training realized from 40 ACM engagements in the simulator does, in fact, improve performance.

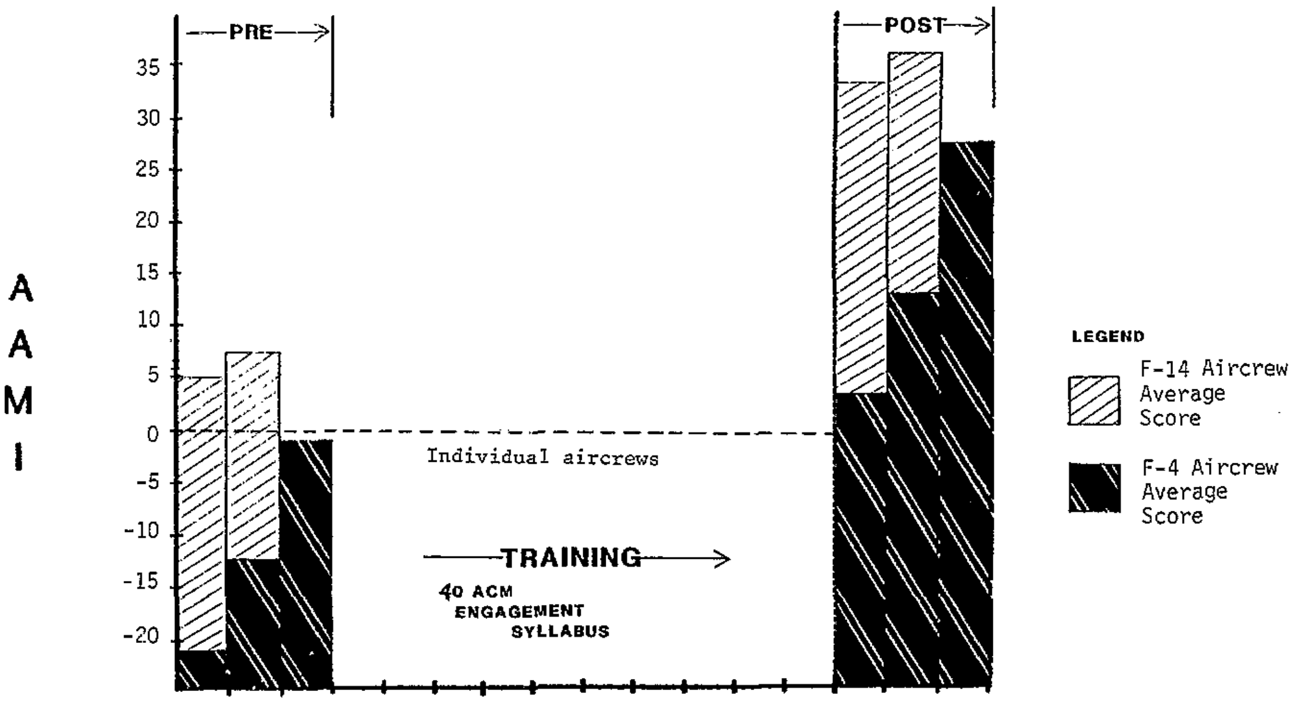
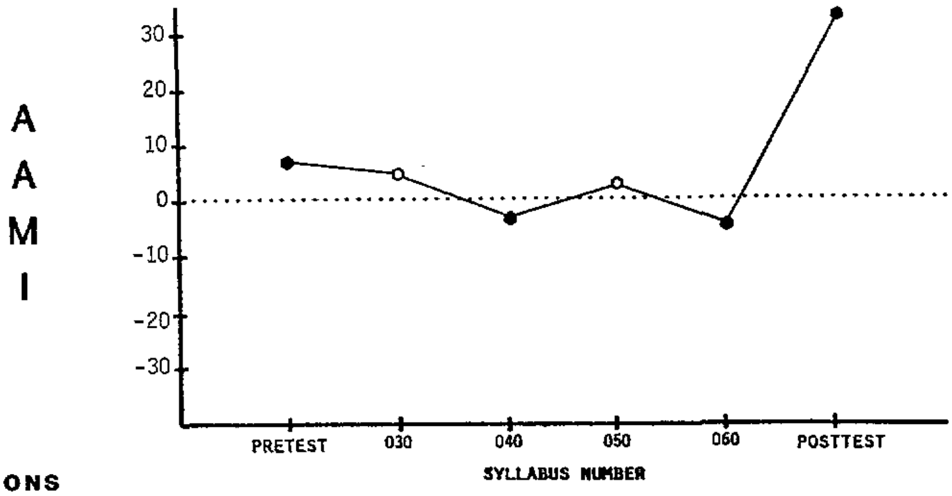


Figure 2. AIR COMBAT MANEUVERING SIMULATOR (ACMS) DATA

Another analysis of F-14 simulator data from the friendly aircraft point of view is contained in Figure 3. The Pre-test data point is an average from seven aircrews over the three engagements comprising the Pre-test. The run conditions and number of engagements experienced were identical for both the Pre-test and Post-test. All simulator runs were stopped when they exceeded three minutes in duration or when 2 successful missile kills were achieved by either the friendly or the bogey. Of significance, Figure 3 demonstrates a substantial increase in performance from Pre- to Post-test. The syllabus numbers in between the Pre- and Post-test represent incremental variations in the training steps applied to the participating aircrews. Each data point in the graph represents the average performance over ten engagements for each syllabus step. The number of aircrews participating, as well as the total number of engagements in each syllabus step, varied across the syllabus due to uncontrollable factors such as aircrew availability and/or simulator equipment problems. The forty engagements experienced between the Pre- and Post-test varied in the level of difficulty presented to the friendly aircraft. As shown in the "run conditions" at the bottom of Figure 3, the first ten engagements in the syllabus (Step 030) had the friendly aircraft loaded with a rear hemisphere missile against a "medium" level of difficulty. A dotted line indicates that no change occured from the previous condition in that syllabus block. Note that a decrease in offensive advantage was registered between the Pre-test and Step 030, during which the friendly fighter lost his 9L all-aspect missile. In addition, during the next ten engagements (Step 040) the adversary difficulty level was increased to "high", which resulted in a drop in the aircrew AAMI average to below zero. That is, moving from a "medium" to a "high" difficulty level opponent resulted in average aircrew performance dropping from an offensive to a defensive posture.

Further analysis of F-14 simulator data contained in Figure 3 demonstrates a substantial increase in performance from Pre-test to Post-test. Switching from "high" difficulty level to "medium" difficulty level when the friendly and adversary aircraft were loaded with similar weapons on the same type of aircraft (both F-14s) resulted in an average offensive advantage for the friendly aircrew as shown by the data point in Step 050. Increasing the level of difficulty from "medium" to "high" while maintaining the same aircraft and weapon relationships resulted in a decrease in average performance below the offensive (zero) score line, once again as shown by the data point for Step 060.

The F-4 data depicted in Figure 4 again indicate that the AAMI reflected relatively high offensive performance in response to an adversary at a "medium" difficulty level when contrasted to performance against a "high" difficulty level. For example, a lower adversary difficulty level at syllabus Step 010, in comparison to the Pre-test, is associated with an increase in the average offensive score. When the high adversary difficulty level is introduced at syllabus Step 031, a performance decrement is noted. The data from the last two syllabus steps and the Post-test data suggest that a learning asymptote may have occurred. Six aircrews participated in the Pre-test; although again the number of aircrews as well as the number of engagements in each syllabus step fluctuated due to uncontrollable factors. As with the F-14 data in Figure 3, the Post-test scores, considered in relationship to the average scores and corresponding standard deviations achieved in other syllabus steps and the Pre-test, suggest that learning has taken place.



RUN CONDITIONS

FRIENDLY WEAPON LOAD	AIM-9L	AIM-9G	--------	AIM-9L	--------	--------
ADVERSARY TYPE	THREAT 1	--------	--------	F-14	--------	THREAT 1
ADVERSARY WEAPON LOAD	AIM-9G	--------	--------	AIM-9L	--------	AIM-9G
ADVERSARY DIFFICULTY LEVEL	HIGH	MEDIUM	HIGH	MEDIUM	HIGH	--------

Figure 3. F-14 SIMULATOR DATA

RUN CONDITIONS

FRIENDLY WEAPON LOAD	AIM-9L	--------	--------	--------	--------	--------
ADVERSARY TYPE	THREAT 1	--------	--------	--------	--------	--------
ADVERSARY WEAPON LOAD	AIM-9G	--------	--------	--------	--------	--------
ADVERSARY DIFFICULTY LEVEL	HIGH	MEDIUM	--------	HIGH	--------	--------

Figure 4. F-4 SIMULATOR DATA

An examination of a single F-4 aircrew is detailed in Figure 5 as an example of the type of data available from the ACM Performance Measurement System (PMS). The average score is a composite measure which can be easily broken into other types and combinations of metric indicators. In this example, the Pre-test average score is -22.4, while the Post-test performance average increases to +12.29. At the same time, the average engagement length decreases markedly from 184 seconds to 82 seconds from Pre-Test to Post-test. The Pre-test also demonstrates another interesting fact. In the beginning of the session, the friendly fighter lost once but never won a Pre-test engagement. At the end of the simulator training, the fighter won all three Post-test engagements. The initial simulator learning period (Step 010, the first ten engagements after the Pre-test) also reflects the fact that the aircrew was learning the syllabus objectives. As demonstrated in Figure 5, the last five engagements of each syllabus Step (010(2) and 040(2)) resulted in the fighter increasing his average offensive score, decreasing the average engagement length and winning four out of five engagements (one engagement was dropped because of computer problems). The same learning trend is illustrated by the data from the concluding simulator learning period also shown in Figure 5.

DISCUSSION

The data presented in this paper are preliminary in nature. The point of this paper is to bring out the fact that a common, objective measurement tool has been developed for ACM, and that the assumptions underlying the system have been validated via actual aircrew training data.

The ACM performance measurement tool can objectively determine differences in:

- Aircraft Performance (Friendly and Bogey)
- Weapons System (AIM-9L, -9G, -7F, Guns)
- Aircrew (Type and Background)
- Tactics
- Training (Equipment, Syllabi)

Examination of the data has been extensive, ranging from initial scatterplots of relevant individual performance points, to variance and correlational analyses to determine the significance of grouped data points and interactions. This examination process, including review by subject matter experts, has verified the validity and reliability of the system as well as the fundamental assumptions of the PMS.

A difficult question is "Does the data collected and analyzed thus far indicate that the air-

			PERFORMANCE AVERAGE		ENGAGEMENTS OUTCOME		
	SYLLABUS STEP	LEVEL OF DIFFICULTY	AAMI SCORE	ENGAGEMENT LENGTH	FIGHTER WINS	THREAT WINS	FIGHTER ESCAPES
PRE TEST	PRE	High	-22.4	184 secs	0	1	0
POST TEST	POST	High	12.29	82 secs	3	0	0
INITIAL SIMULATOR LEARNING	010_1	MEDIUM	13.87	88 secs	1	1	0
	010_2	MEDIUM	30.45	64 secs	4	0	0
CONCLUDING SIMULATOR LEARNING	040_1	High	.71	73 secs	2	1	0
	040_2	High	25.36	68 secs	4	0	0

Figure 5. SIMULATOR DATA F-4 vs. THREAT 1 (example-AIRCREW #18)

crews are learning to fly the simulator?". The answer is that they are, indeed, learning something. The scores improve and the variance decreases in accordance with pertinent syllabus steps and in response to modification in training conditions. Clearly there are "tricks" that can be learned to permit one experienced simulator flyer to beat another simulator aircrew. And, given enough time, some of these tricks may possibly be adapted. For example, it is possible for an aircrew to enhance conditions under which the computer-driven bogey becomes increasingly vulnerable, such as by flying at very high angles of attack or below 200 feet in altitude. However, PSI personnel with ACM experience have been operating and "flying" the simulator for over two years during and before this study. In our opinion, forty simulator engagements, conducted under varying syllabus conditions, does not permit an aircrew enough time to "learn" to trick the simulator. Rather, the data collected thus far and the feedback from the aircrews themselves strongly suggest that more substantive learning is taking place.

A more difficult question is "Do the patterns learned in the ACM simulator transfer to the TACTS range?". The answer to this question is being vigorously pursued. All data collection efforts are being conducted on a not-to-interfere basis. It is difficult to carry out an experimental paradigm on a simulator that is being used for operational training. It is decidely more difficult to apply experimental rigor in the TACTS range environment. Experimental design techniques and statistical controls are being employed to compensate where possible.

Data are presently being collected on the TACTS range as well as from the ACMS to answer both questions. These questions have strong implications for R&D investigations, for training system design efforts, and for combat readiness issues.

ABOUT THE AUTHORS

DR. JAMES McGUINNESS is President/Technical Director of Person-System Integration, Limited. PSI is involved in varied hardware, software and measurement projects related to simulator and training devices. Dr. McGuinness has worked in simulator R&D for McDonnell Douglas and General Dynamics. He holds a Ph.D. from Texas Christian University.

MR. JOHN H. "DUTCH" BOUWMAN is the PSI Virginia Beach Branch Director. As an ex-Navy fighter aircrewman, he has had extensive experience with both the TACTS and ACMS. He holds a Bachelors degree from Northwestern College and is currently pursuing a Masters degree at Old Dominion University.

MR. JOSEPH A. PUIG is a Research Psychologist in the Human Factors Laboratory at the Naval Training Equipment Center. He is currently engaged in the evaluation of training systems and studies in the application of human factors technology to simulators. He holds a B.A. and an M.S. degree from New York University and an M.A. degree in experimental psychology from St. Johns University.