Naval Ocean Systems Center
San Diego, CA 92152-5000

Technical Document 2217
November 1991

Tactical Decision Making Under Stress (TADMUS) Program Study

Initial Design

R. H. Riffenburgh

NAVAL OCEAN SYSTEMS CENTER
San Diego, California 92152-5000

J. D. FONTANA, CAPT, USN
Commander

R. T. SHEARER, Acting
Technical Director

ADMINISTRATIVE INFORMATION

The Tactical Decision Making Under Stress (TADMUS) program is a joint program of the Naval Ocean Systems Center (NOSC), San Diego, CA, and the Naval Training Systems Center (NTSC), Orlando, Fl, under sponsorship of the Office of Naval Technology (ONT), Arlington, VA. This report describes the experiment for NOSC to satisfy Task 1, including Performance Standards, Measures of Performance, administration of experiments, data reduction, and Measures of Effectiveness, with enough background summary to make clear the setting of these aspects. Much of NTSC's Task 2 could follow the same approach if NTSC should wish to use it. Tasks 3, 4, and 5 cannot be definitively designed until at least preliminary results of Task 1 are available. This report represents work conducted during FY 91 with a cutoff date of 30 Sep 91.

Released by
J. A. Howe, Acting Head
Systems Analysis Group

Under authority of
J. T. Avery, Acting Head
Planning, Intelligence, and
Analysis Office

LH

CONTENTS

TABLES

1. INTRODUCTION

The Tactical Decision Making Under Stress (TADMUS) program has begun jointly at the Naval Ocean Systems Center (NOSC), San Diego CA, and the Naval Training Systems Center (NTSC), Orlando FL, sponsored by the Office of Naval Technology (ONT), Arlington VA. NOSC's portion is designated RS34D60.

TADMUS is aimed at the development of aids for decision making in low-intensity conflict (LIC). Making tactical decisions during conflict is by its nature stressful. Thus, the intent of the program is to aid decision making in situations that happen to be stressful, rather than to reduce the stress, which is sometimes misunderstood to be the intent. Products of the program (from both NOSC and NTSC) are to include the following: (1) a body of knowledge to support decision aid development for some of the more stressful LIC tactical situations, e.g., Aegis anti-air warfare (AAW) in LIC; (2) a set of principles to guide LIC decision aid development, including decision support, training, simulation, and display principles; and (3) a laboratory facility to assist the development of LIC decision aids, to be known as the Decision Evaluation Facility for Tactical Teams (DEFTT).

The ONT Program Plan[1] lists five tasks in the study. As the program has evolved, Task 1 falls primarily to NOSC and Task 2 to NTSC. While this paper formally addresses only Task 1, NTSC could follow the same approach for much of Task 2 should it wish. Indeed, the project would benefit from the two laboratories using a common study design. Tasks 3, 4, and 5 depend too much on the outcomes of Tasks 1 and 2 to be addressed in any definitive form at this time.

Task 1 as given in reference 1 is composed of Task Definition and Measurement, which includes developing (1) scenarios, (2) a prototype DEFTT, (3) a performance measurement protocol, and (4) a pilot (hereafter "baseline") experiment to provide baseline data for further experiments. It should be noted that NTSC's Task 2, Examination of Stress Effects on Decision Making, includes selecting stressors (which must be coördinated carefully with item (1) above, as the stressor must fit in the scenario), quantifying their effects (the application of item (3) above), and repeating the baseline experiment with stressors.

The design is intended to provide an overview of the quantitative aspects of Task 1, including the questions needed to be answered; a sketch of the scenario; a statement of the experimental hypotheses; the development of Performance Standards, Measures of Performance (MOPs), the preparation of data, and Measures of Effectiveness (MOEs); design of the experimental analysis; and data analysis.

The following abbreviations, consistent with rather common usage, will be adopted: *DM*: decision maker; *DMg*: decision making.

2. PERSONNEL TEAMS

Personnel involved at NOSC will fall into five de facto groups, whether or not they are formally constituted: (1) the NOSC Technical Team (developers composed of NOSC scientists), (2) the DEFTT Team (DEFTT operators composed of Navy officers/enlisted and NOSC scientists/technicians), (3) a committee composed of experienced naval operators who assign weightings to MOEs, (4) contractors as required, and (5) the Subject Group (composed of Navy officers/enlisted). These teams are not necessarily mutually exclusive; for example, some Technical Team members are likely to be also on the DEFTT Team.

---

[1] Office of Naval Technology. 1990. **FY 90 Program Plan for Tactical Decision-Making Under Stress (TADMUS)**, Arlington VA.

The team distinctions posed here are for administrative clarity and assignment of responsibility. The conduct of the study would be enhanced if these teams were to be formalized and tasked.

3. GOALS

This section lists questions Task 1 should answer.

"The objective of the TADMUS program is to apply recent developments in decision theory, individual and team training, and information display to the problem of enhancing tactical decision quality under conditions of stress" (see footnote 1). The Task 1 steps to achieve this objective are as follows:

Task 1 Step 1. Understand the decision task. The context for this understanding is AAW operations aboard a major warship. Understanding is obtained through studying and analyzing DMg experience as revealed in relevant documents and reports, field observations, interviews with operating personnel.

Task 1 Step 2. Establish laboratory test facility. DEFTT is to be developed with identical facilities at NOSC and NTSC, and possibly elsewhere if required. DEFTT is to include the scenario and its presentation as part of a simulation of an operational decision event along with performance measurement tools, e.g., automated data recording, time-stamped videotaping, and data channel multiplexing.

Task 1 Step 3. Develop measures of effectiveness. Various potential measures of effectiveness (MOEs) are to be examined and prototypes selected. These prototypes should be capable of measuring both baseline DMg performance and DMg performance under experimental influences. Trials are to be conducted and the MOEs refined and improved as required. The MOEs should include both measures of DMg "processes," e.g., quality of reasoning, team coordination, etc.; and of DMg "outcomes," e.g., decision accuracy, latency, consequences, etc.

Task 1 Step 4. Establish baseline decision-making performance. Conduct experiments using Navy Combat Information Center (CIC) teams to provide a baseline performance against which later experiments and variations can be contrasted. These initial experiments will also provide data for the assessment of the scenario, the MOEs, and DEFTT and its operation.

Task 2 continues the objective and should be noted for context. The ONT Program Plan (see footnote 1) gives its steps as follows: Step 1. Understand combat stress (Describe stress aspects of combat DMg and propose a set of stressors--conditions/events which increase the stress in combat DMg); Step 2. Develop stress-inducing methods (Conduct trials with and select the potential stressors to be used in TADMUS); Step 3. Develop techniques to measure stress (Establish numerically measured levels of these stressors to use in experimentation and develop a measurement scheme to quantify experimental effects of the stressors); and Step 4. Establish baseline of decision making under stress (Conduct trials to provide the same sort of baseline that resulted from Task 1 Step 4 for the various stressors).

4. SCENARIO

Fundamental requirements for the scenario were as follows: (1) operation in shallow/confined waters, (2) neutral and hostile countries in close proximity, (3) modern blue/gray systems and weapons among neutral, friendly, and hostile nations, and (4) heavy neutral/friendly traffic in vicinity.

The scenario[2] is set in the Middle East with a mixture of hostile nations, neutral nations, and friendly nations in the vicinity, where several of the nations initially so classed have reason for suddenly changing their loyalties. Thus, the continued applica-

---

[2] Rogers, Will. 10 Sep 91. **TADMUS Scenario Script**, Orincon Corporation, San Diego, California.

bility of the rules of engagement (ROEs) is uncertain, and the intent of an approaching threat from any one nation is uncertain. Also, the wide distribution of blue/gray equipment causes uncertainty as to the national origin of a contact.

The scenario is composed of background information, mission assignment, and a sequence of nine decision situations, or vignettes. Three major uncertainties occur in the vignettes: ROE interpretation, contact identification, and contact intent. Each vignette follows the pattern: Set in a poorly defined situation one or more threats of uncertain origin and uncertain intent approach either own ship or ship being protected and do not respond properly to warnings. The CIC team must decide on a sequence of responses as the situation evolves.

As an example, consider the first vignette. Own ship is escorting the USS *La Salle* (AGF-3) in the Persian Gulf. An air contact emerges from the radar shadow of Iran's central mountains (Point 1). (These points will be referred to below.) It is tracked at 8000 feet following an erratic northwesterly course that will take it to within about 5 nmi of the *La Salle*. It does not respond to challenges, but an air distress signal is intercepted in which the pilot claims to be an Iraqi pilot escaping Iranian internment (Point 2). At 32 nmi, 400 kts, he has descended to 5000 feet and is heading toward the *La Salle* (Point 3: last moment to react).

The vignettes were designed to be unlinked for experimental control and statistical independence with the intent that a decision in one vignette would not influence decisions in later vignettes.

5. EXPERIMENTAL HYPOTHESES

In general, legitimately designed experiments require specifically stated hypotheses to test. This section will provide appropriate hypotheses to be tested in Task 1 Step 4 and perhaps even in Task 2 and later experiments. While the precise wording required to conduct the experiment has not been agreed upon, the hypotheses can be given in some generality. The intent of the TADMUS study is to investigate team DMg. However, no methodology has been published for quantitatively assessing the contribution of team members in a team decision, which forces the TADMUS study to treat the team as an entity for DMg purposes. The rudiments of a methodology for assessing team member contribution have been developed by the author but are as yet unproved. This approach is included as a portion of the analysis in Section 9 below on Conduct and Analysis of the Experiment. Prior to this section, the reader may think of DM as an entity: the CIC team in total.

A team's one full run-through of the scenario vignettes will be called a "game."

Hypothesis 1. DM understands the mission. The "mission" will be stated as a list of mission components of varying priority (e.g., protect escorted ship, protect own ship, do not endanger U.S. political mission in area, ...). "Understand" implies DM shares the same list with the same priority values, expressed as importance ratings, as the Performance Standard, which represents "command authority." The mission remains the same during the game and needs to be measured only at the beginning.

Hypothesis 2. DM adequately assesses the situation. The "situation" is the collection of tactical data and the implication of this collection in terms of the mission. "Assessment" is DM's evaluation of this situation, where "value" implies quantification. "adequately" implies that DM's quantified assessment agrees with the Performance Standard. As the situation evolves, it must be reassessed. DM's situation assessment must be measured at key points, as in the vignette exemplified above.

Hypothesis 3. DM chooses adequate actions to take. "Actions" are tactical steps, e.g., track aircraft at Point 1, pursue identification and prepare air defenses aboard ship at Point 2, and shoot or not at Point 3. "Adequate" implies that the actions chosen agree with the Performance Standard. DM's action choice must be measured at each key point.

6. DATA

The method of quantifying the subjects' experimental behavior and the underlying numerical framework has been developed by Dr. Lawrence Fogel[3] using a variation of a maximum expected utility approach he terms Valuated State Space (VSS). Essentially, the team of very experienced officers assigns numerical relative *importances* in the context of the tactical situation to the various decision opportunities and further assigns numerical relative *values* of the tactical outcomes to the various possible decisions themselves that can be taken at each opportunity. Importance times value yields a tactical *worth* for each of the various decisions, including assessments and act-choices. This (comprehensive) list of worths provides numerical scores for each decision made by the subject DM. This report will not present further details of this method and framework, since that will appear in an update to reference 7, but from here will assume that it exists and it will concern itself with using the "worth"-of-a-decision quantities emerging from Dr. Fogel's VSS. (Dr. Fogel has also contributed to the Performance Standards and MOPs.)

It is assumed that a criterion, including a set of criterion values, from which Performance Standards and MOPs can be obtained, will have been established by running a team of experienced officers through the scenario in the DEFTT, stopping at each decision point for a discussion leading to a consensus of values to be used. This should take about 1 day if full preparations are made. The intent is not to establish what decisions are "right", but what decisions are typical of trained, experienced officers.

The remaining challenge is to identify exactly what the decisions were and when they occurred during the games. Observations will be made during the games, composed of raw data bearing on the hypotheses to be tested. Since these raw data are not in the form required for calculating performance measures, they must be interpreted. Methods to obtain the data on subject's decisions are suggested below. These methods are untried and will doubtless have to be refined.

Hypothesis 1. Observations will be answers to queries that are directly and easily scorable. The scores enter the performance measure.

Hypothesis 2. At key points, data taken will consist of verbal instructions, key strokes, or answers to "the admiral's" queries. From these, DM's choice from the list of possible situations must be inferred. A correspondence key must be prepared to relate the possible raw data to the situations list in order to convert the raw data to situation selection. The situation selection (singular) enters the performance measure.

Hypothesis 3. At key points, data taken will consist of verbal orders or key strokes. From these, DM's choices from the list of tactical actions must be inferred. A correspondence key must be prepared to relate the possible raw data to the possible actions in order to convert the raw data to action selection. The action selections (plural) enter the performance measure.

7. CONCEPTS OF QUANTIFIED MEASURES

Performance Standards describe quantitatively what action *should* be taken by an experienced CIC team under normal combat stress. MOPs describe quantitatively what action actually *was* taken. MOEs relate MOPs to Performance Standards, i.e., effectiveness is shown by contrasting actual performance to performance targeted. Section (7) addresses concepts for these measures.

In the ensuing section (8), formulas for calculating Performance Standards, MOPs, and MOEs are developed. Section 8 may be omitted without conceptual loss by those who do not need to involve themselves in the mathematics.

---

[3] Fogel, Lawrence. 16 May 91. **Draft: Preliminary Experiment Data Acquisition and Analysis**, Orincon Corporation, San Diego, California.

Table 1. Symbols representing Performance Standards, MOPs, and MOEs for 3 hypotheses.

	Performance Standards	Measures of Performance	Measures of Effectiveness
Hypothesis 1	s_1	p_1	e_1
Hypothesis 2	s_2	p_2	e_2
Hypothesis 3	s_3	p_3	e_3

The three measures for each of three hypotheses form nine summary statistics. To keep them straight, a mnemonic device will be used: s will denote standard; p, performance; and e, effectiveness. The hypothesis number will appear as a subscript. Thus, p_2 represents a MOP for hypothesis 2, etc. Table 1 displays this organization.

Hypothesis 1. The criterion for Hypothesis 1 is a quantified list of possible mission components, where quantified implies that each mission component has been assigned an importance weight and each pair: <mission component–importance weight> has an associated worth, i.e., cost–or–benefit of assigning that importance to that component. In terms of Fogel's VSS approach, these worths represent *the importance–of–the–choice weight times the value of having made that choice*. The criterion, or targeted, values form an array, or matrix, with these worths composing the body of the table depending on mission components and importance ratings. These values would appear as listed in table 2.

The Performance Standard for Hypothesis 1, denoted s_1, is the maximum possible score, obtained by summing the largest elements per row of table 2.

Subjects' DMg data for Hypothesis 1 will be collected by inquiry by a DEFTT team member at the close of the initial brief, just before the play (at vignette 1) begins. The DM will be asked for his rating of the importance of each possible mission component on a score of 0 to 10. In table 2, mission component 2 is the most important. If DM rates it so, he scores a 10. If he rates it as importance 9, he scores a 7. The lower he rates it, the lower is his score. If he rates it not important

Table 2. Worths of mission component importance ratings.

		Importance Rating 10	9	8	7	...	0
Mission Components	1	8	10	6	2		−8
	2	10	7	2	−1		−9
	3	−20	−8	−4	−2		10
	.	.	.	.	.	. . .	.

at all, he gets a negative score. Mission component 1 should not be rated most important and DM gets less than full marks, although a positive score, if he rates it so. Mission component 3 is not important at all and should be rated so; DM gets a negative score if he proposes to use his assets to accomplish this undesirable component.

The MOP for Hypothesis 1 for DM k, denoted p_k, will be the sum of his ratings for the mission components.

Given the Performance Standard and MOP, the MOE of DM k for Hypothesis 1, denoted e_{1k}, is simply the percent the observed performance is of the target performance.

Hypothesis 2. The performance standard for Hypothesis 2 is a list, occurring at each key point of each vignette, of possible interpretations of the tactical situation, assigned weights by consensus of the experienced, target-setting team. These interpretations are for overall situations, e.g., "Contact is a threat," rather than for aspects of a situation, e.g., "Contact is close and descending." Each situation interpretation has an associated worth (cost, benefit) of correct or erroneous selection. In VSS terms, these worths again represent the correctness of the choice weight times the value of having made that choice. These worths, for each key point in the scenario, would appear somewhat as listed in table 3.

Table 3. Worths of situation interpretations at a key point.

Situation Interpretation	Worth
1	10
2	2
3	−20
.	.
.	.

The Performance Standard at a key point is the largest worth in the list of worths for that key point, as would be 10 for the key point shown in table 3. The overall Performance Standard for Hypothesis 2, denoted s_2, is the sum of key-point Performance Standards added over the key points.

Subject DMg data for Hypothesis 2 will be taken at each key point. Data for Hypothesis 2 at key point j consists of DM's selection of one from the list of possible situation interpretations. The indicator of DM's situation selection must come from different sources. At many key points, the selection will be clear from the tactical orders given. A careful examination and perhaps some pilot runs must be made to identify the cases where this is not possible. It is possible to have the admiral in command (an actor) ring the CIC team and ask how it perceives the situation, but this cannot be done more than two or three times during the entire game. If these two mechanisms do not exhaust the measurement requirements, further steps must be found, yet unknown.

The performance for a DM observed at a particular key point will be the worth associated with his selection of the situation assessment at that key point. The overall MOP for that DM will be the sum over key points of differences between key-point performance standard and key-point performance observed.

The MOE for Hypothesis 2, i.e., that DM appropriately assesses the situations, is the percent ratio of sum over key points of DM's success (best total assessment score less DM's total of deviations from best assessment scores) to sum over key points of best scores.

Hypothesis 3. The criterion for Hypothesis 3 is much like that for Hypothesis 2: a list, for each key point of each vignette, of the q, say, possible actions that could be taken with assigned weights agreeing with how the target-setting team weighted them, resulting in the worth (cost, benefit) of selecting this act. Since some acts are not independent, we must interpret an "action" as a pattern of acts. The worth values associated with each action pattern compose a paired list for each key point of the format shown in table 4.

Table 4. Worths of action patterns at a key point.

Action pattern	Worth
1	10
2	3
3	−14
.	.
.	.

The Performance Standard for Hypothesis 3 must have a different format from that for Hypothesis 2, as more than one action pattern can be chosen. The best performance would be for DM to select all action patterns having positive worth, avoiding others. The Performance Standard at a key point is the sum of positive worths from table 7.4, and the Performance Standard for Hypothesis 2 is the sum of these worth-sums across key points.

The selection of action patterns, i.e., the subject performance data for Hypothesis 3, taken at each key point, will be indicated by the orders DM gives. The performance of DM observed at a key point will be the sum of worths for the action patterns DM chose. The overall MOP for that DM will be the sum over key points of differences between key-point performance standard and key-point performance observed.

The MOE for Hypothesis 3, i.e., that DM takes the right actions, is the percent ratio of sum over key points of DM's success (best action choice score less DM's total of deviation from best action choice scores) to sum over key points of best action choice score.

8. CALCULATION OF PERFORMANCE STANDARDS, MOPs, AND MOEs

(This section may be omitted by those who wish to confine themselves to only the conceptual level.)

8.1 HYPOTHESIS 1

Subscripts:

- i — mission component, i = 1,...,m
- j — importance rating, j = 1,...,n
- k — DM (subject designator)

Worth for i^{th} mission component with j^{th} importance rating:

s_{1ij} — entry from table 2, row i, column j

Performance Standard for Hypothesis 1:

$$s_1 = \sum_{i=1}^{m} \max_j (s_{1ij})$$

Subject performance component for i^{th} mission component:

p_{1ik} — Rating by k^{th} DM of importance

MOP for k^{th} DM for Hypothesis 1:

$$p_{1k} = \sum_{i=1}^{m} p_{1ik}$$

MOE for k^{th} DM for Hypothesis 1:

$$c_{1k} = 100\, p_{1k} / s_1 \ .$$

8.2 HYPOTHESIS 2

Subscripts:

- i — situation interpretation, i = 1,...,m
- j — key point in scenario vignette, j = 1,...,n
- k — DM (subject designator)

Worth for i^{th} situation interpretation at j^{th} key point:

s_{2ij} entry from j^{th} issue of table 3

Performance Standard at j^{th} key point:

$$s_{2j} = \max_{i} (s_{2ij})$$

Performance Standard for Hypothesis 2:

$$s_2 = \sum_{j=1}^{n} s_{2j}$$

Subject performance component for k^{th} DM at j^{th} key point:

p_{2ik} Worth of DM's choice of situation assessment from j^{th} table 3

MOP for k^{th} DM for Hypothesis 2:

$$p_{2k} = \sum_{j=1}^{n} (s_{2j} - p_{2jk})$$

MOE for k^{th} DM for Hypothesis 2:

$$c_{2k} = 100 (s_2 - p_{2k}) / p_2$$

8.3 HYPOTHESIS 3

Subscripts:

- i — action pattern, i = 1,...,m
- j — key point in scenario vignette, j = 1,...,n
- k — DM (subject designator)

Worth for i^{th} action pattern at j^{th} key point:

s_{3ij} entry from j^{th} issue of Table 4

To select patterns with only positive worth, let us define a symbol δ_{ij} such that

$$\delta_{ij} \begin{cases} = 1, & s_{3ij} > 0 \\ = 0, & s_{3ij} \leq 0 \end{cases}$$

Performance Standard at j^{th} key point:

$$s_{3j} = \sum_{i=1}^{m} \delta_{ij} s_{3ij}$$

Performance Standard for Hypothesis 3:

$$s_3 = \sum_{j=1}^{n} s_{3j}$$

To sum only actions selected by DM, let us define a symbol c_{ijk} such that

$$c_{ijk} \begin{cases} = 1, & \text{action i selected at key point j by DM k} \\ = 0, & \text{action i not selected at key point j by DM k} \end{cases}$$

Performance component for DM k at key point j (sum of utilities of actions he chooses):

$$p_{3jk} = \Sigma_i c_{ijk} s_{3ij}$$

MOP for DM k for Hypothesis 2:

$$p_{3k} = \sum_{j=1}^{n} (s_{3j} - p_{3jk})$$

MOE for DM k for Hypothesis 1:

$$e_{3k} = 100 (s_3 - p_{3k}) / s_3 \ .$$

9. CONDUCT AND ANALYSIS OF THE EXPERIMENT

9.1 SAMPLE SIZE, PER CIC TEAM

Nine vignettes per game are planned. The relationship of these nine to sample size is one of independence. The DMg difficulty will vary by vignette, as the vignettes are precluded from being standardized experimentally by the lack of sample availability. Another consideration is the effect of early action taken on later decisions. Although every effort is being taken to prevent such effect, primarily by witholding outcome information from DM during the game, it is still possible that a decision to fire at, say, an Iranian aircraft in an early vignette may affect DM's decision to fire again or not at another Iranian aircraft in a later vignette.

If we may assume that the sequence of nine vignettes is independent one from the other, we can consider each game, i.e., the run-through of the scenario by a CIC team, to have a sample size of one for Hypothesis 1 and sample size nine for Hypotheses 2 and 3. This assumption implies that the measures of DM's situation assessment and selection of tactical actions are not influenced by (1) difficulty within vignette, (2) tactical action taken early in the game, (3) learning between the first and last vignettes presented, and (4) any loss in reality or player seriousness between the first and last vignettes. The independence assumption would be strengthened by randomizing the order of presentation of vignettes.

9.2 SAMPLE SIZE, BETWEEN CIC TEAMS

It is believed that five or six Aegis teams will be available over a several-month period. Non-Aegis CIC teams may be used and will be tapped, but some changes in scenario, DEFTT, and measurement process will be required. For example, the Combat Systems Coordinator position must be removed. Inasmuch as Tasks 3, 4, and 5 will require subject teams and 2 years or so must elapse before enough changes have occurred for re-use of the CIC team from a given ship, it is anticipated that samples will continue to be in short supply. For the moment, let us assume that non-Aegis teams will be used for later tasks and that we will be constrained to six teams for Task 1 (and, probably, Task 2).

These six teams will be different one from another, due to team members' varying experience and personalities (dominance, communication habits, etc.). The extent of this very difference is an interesting question that must be investigated by the study. If not included in the study design, this factor could have a confounding effect on the statistical design. The way to incorporate this factor into the statistical analysis can not be planned until information about the natures and differences of the teams is at hand. Thus, the between-team sample size may be six, or three, or two, or even one, and this will not be known until the data have been taken.

9.3 MEASURES ON TEAM AS A WHOLE VS. INDIVIDUAL TEAM MEMBERS

The data and analysis in the decision has been treated so far in this design as a team output. However, it is also of interest to examine the DMg character of individuals on the teams to learn (1) what relative contribution individual team members make to team decisions, (2) how team member quality (superior or inferior) affects the quality of team decisions, and (3) how communication patterns (varying personalities and position dominance) affect the team decisions.

Fogel's technique breaks the steps for collecting information and recommending action into components assignable to the DM, which has been taken so far as an entity. Without too much additional effort, we may record the contribution of each team member, from which we may calculate the influence on the decision made by each team member. With this information, we can subject both whole-team data and individual-team-member data to our analysis. A comparison of team vs. individual statistical results can address items (1) and (2) in the paragraph above. However, item (3) must be done in close connection with a psychologist and is not planned for TADMUS.

An example of individual-contribution assessment may be useful. An Aegis AAW team consists of seven members: Commanding Officer (CO), Tactical Action Officer (TAO), Combat Systems Coordinator (CSC), Tactical Information Coordinator (TIC), Antiair Warfare Coordinator (AAWC), Identification Supervisor (IDS), and Electronic Warfare Supervisor (EWS). Fogel has listed the functions of the team in responding to a threat and obtained experienced-officer ratings of the relative importance of each task, and then combined these importances numerically to provide overall importance. Table 5 lists a few of these many tasks as an example. The first column to the right of the list shows the relative importance ratings by function for the team as a whole. Then the team members are listed, with the proportion contribution each makes to that function. The product of function importance times member responsibility yields a measure of member contribution per function.

So far we have discussed obtaining data on contributions the team members *should* make and the contributions they *do* make. Additionally, asking the team members individually to fill in the team-member contribution table after the game would give the perception of each member of the distribution of responsibilities. These three sets would

Table 5. Some AAW-team tasks and associated relative importances of team decisions, with team members and proportion contribution to each task made by each team member

Some tasks	Team decision relative importance	Team Members						
		CO	TAO	CSC	TIC	AAW	IDS	EWS
Assess potential threat								
Closure rate	.034				1.0			
Response to IFF	.031							1.0
Assess current intell	.019	0.2	0.3				0.5	
. . . . .								
Estimate intent of threat								
To attack	.048	1.0						
. . . . .								

allow analysis to uncover actual responsibility, the contrast between perceived and actual responsibility, intra-team variability in responsibility, and between team variability in intra-team responsibility.

9.4 ACCOMPLISHMENT OF TASKS AND STEPS; NTSC INVOLVEMENT

Task 1 Steps 1 and 2 and Task 2 Steps 1 and 2 are not experimental steps. They are developments, being done by NOSC and NTSC respectively. Dr. Fogel's draft (reference 7) and this paper establish Task 1 Step 3. (Task 2 Step 3 is to be done by NTSC.) This paper also plans for the analysis required in Task 1 Step 4 in order to establish an experimental baseline. Task 2 Step 4 is closely related to Task 1 Step 4, apparently following it closely or even being done in conjunction, but cannot be planned by NOSC. It is suggested that, assuming independence among the vignettes and key points, half the key points be free of experimental stressors and half be subject to experimental stressors. The analysis can be redefined with half the observations for each; DMg baselines -- without and with added stressors -- can then be produced. This way the very small sample can be shared between the two Centers, but the detailed and timely participation of NTSC would be required.

9.5 EXPERIMENTAL ANALYSIS

For each hypothesis, there are three characteristics to be assessed to provide a baseline for future experimentation: How the typical baseline team's DMg quality compares with the criterion; how variable baseline teams are one from another; and what probability distribution parameter estimates are for the typical baseline team. These three characteristics will provide the basic quantification required to assess the effects of experimental variables to be used in later TADMUS work. The following analysis method will assume the limited sample size of six teams. It will answer three specifically posed questions, where the subscripts indicating hypotheses 1, 2, and 3 are omitted and can be affixed respectively for each hypothesis. α refers to the probability of a Type I error chosen by the experimenter.

For each hypothesis, the data consist of a percent success measure for each DM (i.e., CIC AAW team), looking something like the following.

Team number (k):	1	2	3	4	5	6
E_k:	93	68	71	84	55	89

Three baseline questions, Q1, Q2, and Q3, are asked. With each is given a statistical method to answer the question.

Q1. Is the average effectiveness of operational teams statistically significantly below the standard for best performance?

Statistical hypotheses:

H_0: population MOE = 100

H_1: population MOE < 100 .

Perform an ordinary *t*-test for significance of difference between performance standard and performance observed. An α = 0.10 is suggested. After the data are collected, if *t* is too insensitive or assumptions seem to be violated, nonparametric or other techniques can be considered.

Q2. Are team MOEs statistically significantly different one from another?

Randomly select two sets of three teams each so that there become two groups. Perform a randomized analysis of variance (ANOV) on the MOEs. A significant F implies that differences do exist. The ANOV table would look like the following.

Source	df	SS	MS	F
Teams	1			
Error	4			
Total	5			

Q3. What are baseline descriptors of team effectiveness which can be used for later comparison?

The historically most useful measures, available from the analyses above, are the mean and standard deviation of effectiveness:

$\bar{c}$, s_c .

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	November 1991	

4. TITLE AND SUBTITLE
TACTICAL DECISION MAKING UNDER STRESS (TADMUS) PROGRAM STUDY
Initial Design

5. FUNDING NUMBERS
PE: 0602233N
WU: DN300090

6. AUTHOR(S)
R. H. Riffenburgh

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Ocean Systems Center
San Diego, CA 92152–5000

8. PERFORMING ORGANIZATION REPORT NUMBER
NOSC TD 2217

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Technology
Arlington, VA 22217

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

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

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

TADMUS, a joint NOSC/Naval Training Systems Center (NTSC) program, is aimed at the development of aids for decision making in low-intensity conflict. This report describes the NOSC experiment to satisfy Task 1, including Performance Standards, Measures of Performance, administration of experiments, data reduction, and Measures of Effectiveness, with enough background summary to make clear the setting of these aspects. Much of NTSC's Task 2 could follow the same approach if NTSC should wish to use it. Tasks 3, 4, and 5 cannot be definitively designed until at least preliminary results of Task 1 are available.

14. SUBJECT TERMS
Antiair Warfare (AAW)
Combat Information Center (CIC)
Decision Making
Measures of Effectiveness (MOE)
Measure of Performance (MOP)
Performance Standard
Tactical Decision Making
Tactical Decision Making under Stress (TADMUS)

15. NUMBER OF PAGES
19

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAME AS REPORT

21a. NAME OF RESPONSIBLE INDIVIDUAL	21b. TELEPHONE *(include Area Code)*	21c. OFFICE SYMBOL
R. H. Riffenburgh	(619) 553–5629	Code 171

INITIAL DISTRIBUTION

Code 0012	Patent Counsel	(1)
Code 0144	R. November	(1)
Code 171	J. T. Avery	(1)
Code 171	G. Ruptier	(1)
Code 171	D. Smith	(1)
Code 171	R. Riffenburgh	(47)
Code 44	J. Grossman	(1)
Code 442	L. Duffy	(1)
Code 442	C. Englund	(1)
Code 442	S. Hutchins	(1)
Code 442	S. Murray	(1)
Code 444	K. Pickslay	(1)
Code 952B	J. Puleo	(1)
Code 961	Archive/Stock	(6)
Code 964B	Library	(3)

Defense Technical Information Center
Alexandria, VA 22304-6145 (4)

Navy Acquisition, Research & Development Information Center (NARDIC)
Alexandria, VA 22333

Naval Military Personnel Command
Washington, DC 20370

Office of Naval Technology
Arlington, VA 22217

George Mason University
Fairfax, VA 22030

University of South Florida
Tampa, FL 33620

Engineering Research Associates
Vienna, VA 22180

Klein Associates, Inc.
Fairborn, OH 45324

Search Technology
Norcross, GA 30092

NOSC Liaison Office
Arlington, VA 22217

Naval Health Research Center
San Diego, CA 92186-5122 (4)

Naval Training Systems Center
Orlando, FL 32826-3224 (5)

Tactical Training Group Pacific
San Diego, CA 92147-5080

Old Dominion University
Norfolk, VA 23529 (2)

CHI Systems, Inc.
Spring House, PA 19477

Florida Maxima Corporation
Winter Park, FL 32789

Orincon Corporation
San Diego, CA 92121 (2)