HAY SYSTEMS, INC.

FINAL REPORT

Generic MANPRINT Analysis

Methodology to Incorporate
Human Factor Variables into
Army Combat Models
(GM/AME-HF)
(Delivery Order Number 0022)

April 1991

Prepared for:

U. S. Army Training
and Doctrine Command
(TRADOC)
ATTN: ATCA
Ft. Eustis, VA 23604-5538

Prepared by:

HAY Systems, Inc.
2000 M Street, N.W.
Suite 650
Washington, DC 20036

Contract Number:
DABT60-87-D-3873

UNCLASSIFIED

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION
Unclassified

1b. RESTRICTIVE MARKINGS
N/A

2a. SECURITY CLASSIFICATION AUTHORITY
N/A

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE
N/A

3. DISTRIBUTION AVAILABILITY OF REPORT

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
A007 For Final Report: Volume 1
A008 for Final Report: Volume 2

5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION
HAY Systems, Inc.

6b. OFFICE SYMBOL (If applicable)

7a. NAME OF MONITORING ORGANIZATION
U.S. Army Training and Doctrine Command

6c. ADDRESS (City, State, and ZIP Code)
2000 M Street, N.W.
Suite 650
Washington, DC 20036-3307

7b. ADDRESS (City, State, and ZIP Code)
ATTN: ATCA
Ft. Eustis, VA 23604-5538

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
USAPIC

8b. OFFICE SYMBOL (If applicable)
ATNC-NMF-C

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
DABT60 87-D-3873

8c. ADDRESS (City, State, and ZIP Code)
ATTN: ATNC-NMF-C
200 Stovall Street
Alexandria, VA 22332-1345

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification) Methodology to Incorporate Human Factor Variables into Army Combat Models

12. PERSONAL AUTHOR(S)
Terry Garrett Joel Schendel Lanny Walker

13a. TYPE OF REPORT
Final

13b. TIME COVERED
FROM 3/90 TO 4/91

14. DATE OF REPORT (Year, Month, Day)
1991 April 30

15. PAGE COUNT
87

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Task Rating Sheet, Unit Systems Effectiveness Variables, Performance Shaping Functions, Performance Prediction Matrix Soldier Performance Variables

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The purpose was to develop a methodology to include the soldier in the Army's combat models. The basic approach entails six steps:

Step 1: Identify candidate combat model and model processes.
Step 2: Identify a set of model unit and systems effectiveness variables that can be influenced by soldier performance.
Step 3: For the protoype effort, select a candidate soldier perfomance variable.
Step 4: Define a method for predicting how soldier performance will be affected by this variable.
Step 5: Establish means for modifying model unit and system effectiveness variable data based on the Step 4 prediction, thereby creating performance shaping functions.
Step 6: Recommend possible approaches for adding the performance shaping functions to a combat model.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
[X] UNCLASSIFIED/UNLIMITED [ ] SAME AS RPT. [ ] DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL
Dr. Phillip Vandivier

22b. TELEPHONE (Include Area Code)
(317) 543-6882

22c. OFFICE SYMBOL
ATRC-B

DD Form 1473, JUN 86 Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

DISCLAIMER

The findings of this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Statement A per telecon
Jean A Bredeson Army
TRADOC Analysis Command/
ATRC-B Ft Benjamin
Harrison, In 46216-5000
NWW 8/22/91

SPECIAL TECHNICAL NOTICE

This paper is an unofficial document intended for wide distribution to obtain comments. This study was performed by HAY Systems, Inc. under contract number DABT60-87-D-3873. Dr. Phillip Vandivier of TRADOC Analysis Command - Ft. Ben Harrison was the contracting officer's representative responsible for the monitoring of this effort. This page, along with the "Responses to Staffing Concerns" found at the end of this paper, are the only two additions made by the TRADOC Analysis Command - Ft. Ben Harrison in sending this paper into publication.

The views, opinions, and/or findings contained herein are those of the authors and should not be construed as the official position of the TRADOC Analysis Command, US Army Training and Doctrine Command, or an official Department of the Army position, policy, or decision unless so designated by other official documentation.

Comments should be sent to:

Director, US Army TRADOC Analysis Command - Ft. Harrison
ATTN: ATRC-B (Bldg 401B)
Ft. Benjamin Harrison, IN 46216-5000

FORWARD

AirLand battlefield doctrine predicts that future wars will require Army units to operate 24 hours per day in continuous or sustained operations as evidenced in Desert Storm. The key limiting factor in these type of stressful operations is, as always, the individual soldier. The Army seeks to predict its future battlefield requirements through the use of combat effectiveness models that includes training and analyses as follows:

- Force structure analysis
- Training and doctrine analysis
- New equipment training assessments
- Weapon systems effectiveness and tradeoff analyses
- Training of commanders and staffs

The trend to use combat modeling in peacetime to prepare for war will continue for the immediate future. One area of modeling that has not been adequately represented is the individual soldier and the systems designed to sustain him.

The purpose of this effort is to identify potential means for including the soldier in combat models. It is not our intent to address the need to account for the soldier in combat models. For example, it has been argued in combat modeling that, in a force-on-force engagement, opposing human factors may "cancel out" or make little difference to the outcome of the battle compared to opposing hardware factors. Therefore, our concern is strictly a methodological one. To achieve this goal, we sought to find an approach that would allow us to model soldier performance without adding substantially to the size or complexity of existing combat models.

We staffed the draft copy of this report with the advisory group and made every attempt to implement their constructive comments that was feasible under the scope and limitations of this project. Member organizations of the advisory group that provided comments are as follows:

- TRADOC Analysis Command
  Ft. Benjamin Harrison, IN

- TRADOC Analysis Command
  White Sands Missile Range, NM

- Model Improvement and Study Management Agency
  Washington, DC

- US Army Research Institute
  Alexandria, VA

- US Army Laboratory Command
  Delphi, MD

- TRADOC Analysis Command
  Monterey, CA

METHODOLOGY TO INCORPORATE HUMAN FACTOR VARIABLES INTO ARMY COMBAT MODELS

CONTENTS

	Page
INTRODUCTION	1
Background	1
Large Number of Human Performance Variables	2
Large Number of Unknowns	3
Limited Amounts of Usable Data	3
Interaction Effects	3
Purpose	4
METHOD	4
Step 1: Identify Candidate Model and Model Processes	5
Step 2: Identify a Set of Model Unit and Systems Effectiveness Variables that can be Influenced by Soldier Performance	7
Step 3: For the Prototype Effort, Select a Candidate Soldier Performance Variable	9
Step 4: Define a Method for Predicting How Soldier Performance will be Affected by the Candidate Variable	9

Step 5: Establish Means for Modifying Model Unit and Systems Effectiveness Variable Data Based on the Step 4 Predictions, Thereby Creating Performance Shaping Functions. 22

Step 6: Recommend Possible Approaches for Adding the Performance Shaping Functions to a Combat Model 27

RESULTS AND DISCUSSION 28

ISSUES AND DIRECTIONS FOR FUTURE RESEARCH 32

REFERENCES 34

Appendices

APPENDIX A. VIC Systems Variables Possibly Influenced by Soldier Performance 39

APPENDIX B. Rating Scale Method 40

APPENDIX C. Sleep Loss Effects Task Rating Sheet 44

APPENDIX D. Sleep Loss Effects Task Rating Sheet References. 49

APPENDIX E. Sleep Loss Performance Data. 58

APPENDIX F. Multiple Regression Analysis. 74

List of Tables

Table 1. VIC Model Processes 6

Table 2. Unit and Systems Effectiveness Variables that can be Influenced by Soldier Performance (VIC Model) 8

Table 3. Sleep Loss Effects Prediction Matrix. 17

Table 4. Analysis of Variance Table. 20

Table 5. Unit and Systems Effectiveness Variable Data -- Target Acquisition Times in Seconds -- VIC Model 23

Table 6. Sleep Loss Effects Task Rating Sheet Scores for Target Acquisition 25

Table 7. Predicted Target Acquisition Times (sec) by Range Interval and Target Status. 26

List of Figures

Figure 1. Method 4

METHODOLOGY TO INCORPORATE HUMAN FACTOR VARIABLES INTO ARMY COMBAT MODELS

INTRODUCTION

Background

Combat models, including simulations and war games, are used by the Army and other military services to support a wide variety of training and analysis activities (e.g., force structure analysis, training and doctrine analysis, weapons systems effectiveness and tradeoff analysis). The advantages of the approach are readily apparent. It affords means to rapidly portray and manipulate various aspects of military operations with far greater control and at far less cost and risk than ever would be possible under everyday, operational conditions. Of course, there is at least one potential drawback to the approach: combat models generate "modelled" results. These results are only as valid as the data and the modeling assumptions on which they are based.

As the Army increases its stake in the combat modeling approach, efforts are being directed toward assessing model results and improving model representations of combat. However, concerns continue to be voiced over the need to enhance the fidelity or realism of these models. Underlying many of these concerns is the general failure of combat models to account for the human aspects of combat.

Combat models today are almost exclusively "firepower," or equipment models. The models were designed to portray the performance characteristics of the equipment, not to consider those of the soldier. As a result, model outputs do not account for the effects of such things as sleep loss, fatigue, temperature extremes, fear, or stress. They assign no value to variables such as combat experience, morale, unit cohesion and esprit, leadership, and training. Only the equipment drives the battle.

As noted by Van Nostrand (1988, p. 6-7):

> Data values for variables such as firing range and probability of hit now usually represent equipment capability, assuming the 'perfect' soldier--each soldier makes no errors in finding all targets at the maximum range, chooses the one with the highest priority for killing, identifies it correctly, instantly makes the correct decision to fire with the correct weapon and ammunition, fires at the maximum weapon range, and with no hesitation chooses the target with the next higher priority, until all targets are killed or he himself is killed. For example, last year one of the analytic combat models at [U.S. Army Concepts Analysis Agency] used probabilities of hit for the M1 tank which ranged up to 3000 meters. Using these probabilities, many targets were killed between 1500 and 3000 meters. Meanwhile data from the National Training Center (NTC) at Fort Irwin, California showed that the great majority of targets were killed at 1500 meters or less. Fewer than 40 percent of the tank platoons had even one tank which fired at ranges of 2000 meters or greater.

The failure of combat models to account for soldier performance and behavior is inconsistent with historical analyses of combat. This inconsistency acts generally to reduce the credibility of model outputs, even though--from a purely hardware perspective--they may be quite accurate. Of course, there are reasons why soldier performance has not been routinely considered in combat models. Some of these reasons are highlighted in the following paragraphs.

Large Number of Human Performance Variables

One reason that soldier performance variables probably have not been included in combat models is the large number of variables that potentially could be considered. In a recent study on the subject, Vandivier (1990) highlighted some 23 different variables as potentially influencing soldier performance. This list is representative of the types of variables that are frequently cited in Department of Defense (DoD) reports, but it is not exhaustive. Numerous other variables have been suggested elsewhere (e.g., Van Nostrand, 1986).

Large Number of Unknowns

Tests of the effects of the same human performance variables do not always yield consistent results or results that appear consistent given other experiential or historical data. Many questions remain to be resolved. Others have been resolved but only after much time-consuming research.

Limited Amounts of Usable Data

A key problem for the modeling community has been the fundamental lack of usable human performance data. Modeling demands for data far outstrip supplies, and the data that do exist vary in terms of quality and relevance. As a result, significant gaps in knowledge exist related to the effects of seemingly critical human performance variables (e.g., Van Nostrand, 1986). This is not to suggest that predictions cannot be made based on data that exist, historical accounts of men in combat, or subjective judgment. However, given the current state of research, not all of these predictions can be expected to be empirically based.

Interaction Effects

It is one thing to predict the effects of some particular variable on performance. It is far more difficult to know how that variable will affect performance when it is treated in combination with other variables. As a simple illustration, Wilkinson (1963) looked at the joint effects of 32 hours of sleep deprivation and intense noise (100 decibels) on performance of a serial choice reaction time task. With normal amounts of sleep, a high level of noise, as might be expected, caused increasing deterioration in performance. Similarly, when subjects performed in a quiet environment but were sleep-deprived, there was an accelerated decrement in performance over time. However, subjects who performed during intense noise and who were sleep-deprived actually had fewer errors than subjects who performed under conditions of sleep deprivation only! Due to interaction effects such as these, modeling the effects of even a small number of soldier performance variables will be difficult.

Purpose

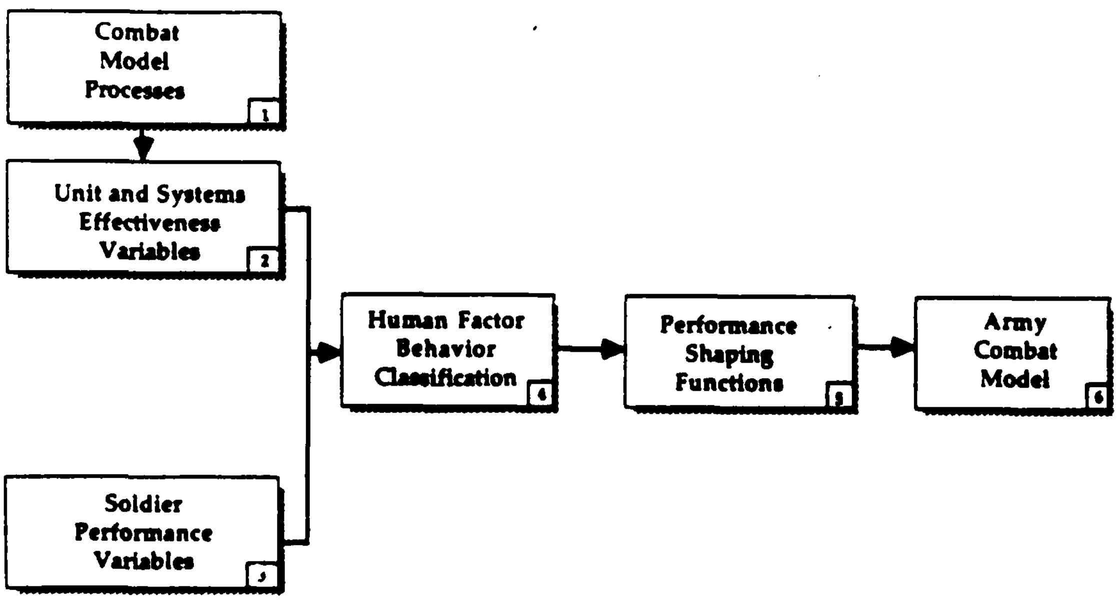
In carrying out this work, it was recognized that numerous questions have been raised about the nee'd to account for the soldier in combat models. For example, it has been argued that, in a force-on-force engagement, opposing human factors may "cancel out" or make little difference compared to opposing hardware factors. It is not our intent to address these issues. Our concern is strictly a methodological one.

The purpose of the present work was to identify potential means for including the soldier in combat models. To achieve this goal, we sought to find answers to issues such as those noted above. In addition, we sought to find an approach that would allow us to model soldier performance without adding substantially to the size or complexity of existing models.

METHOD

Figure 1

Overview of Methodology



Our basic approach to the problem is shown in Figure 1. As suggested by the figure, the approach entails six steps:

- Step 1: Identify candidate combat model and model processes.
- Step 2: Identify a set of model unit and systems effectiveness variables that can be influenced by soldier performance.
- Step 3: For the prototype effort, select a candidate soldier performance variable.
- Step 4: Define a method for predicting how soldier performance will be affected by this variable.
- Step 5: Establish means for modifying model unit and systems effectiveness variable data based on the Step 4 predictions, thereby creating performance shaping functions.
- Step 6: Recommend possible approaches for adding the performance shaping functions to a combat model.

Step 1: Identify Candidate Model and Model Processes

As part of Step 1, we identified a number of potential candidate combat models, to include Vector-In-Commander (VIC) (Department of the Army, 1979a), CARMONETTE, BLDM, and CASTFOREM. Briefly, the VIC model is a deterministic, force-on-force model that reflects combined arms operations at corps level and below. It is used by the Army to assess force structure, new equipment training, and weapon systems acquisition. CARMONETTE is a Monte-Carlo simulation of ground combat generally used to represent combat at company to battalion level. BLDM is a deterministic model of ground combat at company to battalion level. CASTFOREM is the U.S. Army Training and Doctrine Command's (TRADOC's) primary high-resolution simulation of battalion combat.

The VIC combat model was selected initially to illustrate the methodology. This is not to imply that VIC was formally evaluated against the models cited above. VIC was selected because it is representative of the Army's combat model inventory (Department of the Army, 1979a). This inventory currently includes at least 279 combat models (Vandivier, 1990). VIC also uses very specific, highly detailed input values. Typical values may describe the performance of a single weapon system versus a single target. The target description may include type of target (e.g., tank, personnel carrier), status of target (e.g., stationary or moving), and range (meters). A high level of detail was seen as critical to efforts to track the effects of including soldier performance considerations in combat models.

Miller and Bonder (1982) analyzed nine combat simulations to identify human performance interactions providing the 15 VIC model processes presented in Table 1 . These processes may be regarded as steps, actions, or operations used to bring about a desired modeled result. The processes are largely common to both offensive and defensive operations.

Table 1

VIC Model Processes

• Ground Force Deployments	• Combat Service Support
• Command and Control	• Smoke Operations
• Information Processing	• Support Fire Operations
• Intelligence and Fusion Processing	• Helicopter Operations
• Electronic Warfare	• Fixed Wing Air Operations
• Manuever Unit Combat	• Air Defense
• Engineer Operations	• Chemical

Step 2: Identify a Set of Model Unit and Systems Effectiveness Variables that can be Influenced by Soldier Performance

Table 2 shows the total number of variables used to model each of the various model processes. It also shows the total number of variables per model process that appear subject to the effects of soldier performance. These latter values were derived based on a preliminary assessment of the data input variables described in the VIC Data Input and Methodology Manual (Department of the Army, 1979b). The assessment entailed separating those variables that can be influenced by soldier performance from those that are driven strictly by the scenario (e.g., number of red and blue weapon systems) or determined by the engineering characteristics of the weapon systems in use (e.g., target vulnerability).

As an example, maneuver unit combat is a model process that is a representation of fire and maneuver of front-line forces. Ninety-three (93) unit and systems effectiveness variables can be manipulated for maneuver unit combat. Of these variables, sixteen (16) were seen as possibly being influenced by soldier performance. These variables are listed in Appendix A.

Target acquisition and selection was among these variables. It will be used throughout the remainder of the approach section to illustrate the manner in which selected data inputs can be modified to account for soldier performance. The VIC Data Input and Methodology Manual (Department of the Army, 1979b) defines target acquisition and selection as follows:

> ...a target must be acquired and selected before it can be fired on. In order for direct fire target acquisition to occur in the model, line of sight must exist between the observing weapon and its potential target. Line of sight is represented analytically in the module as a function of the type of terrain on which the engagement is occurring and the observer/target range. If line of sight exists, acquisition may occur by either of two target acquisition processes; serial or parallel. Weapons which employ serial acquisition alternately search for and fire at targets, while weapons employing parallel

acquisition can search for new targets while engaging one previously acquired. Finally, the highest priority target acquired then is selected for engagement.

Presumably, any variable that is influenced by soldier performance potentially could have been selected for demonstration purposes.

Table 2

Unit and Systems Effectiveness Variables that can be Influenced by Soldier Performance

Process	Total Number of Variables	Number of Variables that are Influenced by Soldier Performance
Ground Force Deployments	88	20
Command and Control	43	18
Information Processing	76	18
Intelligence/Fusion Processing	113	6
Electronic Warfare	84	10
Maneuver Unit Combat	93	16
Engineer Operations	192	40
Combat Service Support	223	69
Smoke Operations	60	2
Support Fire Operations	174	29
Helicopter Operations	28	10
Fixed Wing Air Operations	200	22
Air Defense	99	7
Chemical	45	14
TOTAL	1518	281

Step 3: For the Prototype Effort, Select a Candidate Soldier Performance Variable

In selecting a candidate soldier performance variable, all of the many variables noted by Vandivier (1990) were considered. At this point, it was decided to focus on sleep loss. Sleep loss was selected for a number of reasons: First and foremost, it is known to be a potent variable for human performance (e.g., Krueger, 1989), and likely to be a key determinant of soldier performance during continuous and sustained operations.

The authors also were predisposed to select a variable, like sleep loss, that has a strong research base and that has been used as a basis for related model developments (e.g., McNally, Machovec, Ellzy, & Hursh, 1989). As will become evident, our methodology does not depend on the presence of a well-developed research base. In fact, it requires no more than subjective judgment. However, a strong research base is desirable and would assure the validity of the predictions that are made.

Step 4: Define a Method for Predicting How Soldier Performance will be Affected by the Candidate Variable

At least two methods were identified for generating specific predictions about the effects of particular variables on soldier performance. Both methods depend on the use of task ratings. The main difference between the methods is in the manner in which these ratings are developed and treated.

The first method depends most directly on the development of a rating instrument, collection of task ratings using this instrument, and correlation of these ratings with observed performance. The result is a prediction matrix, or table, that can then be used to produce some very specific performance predictions. This method has been shown effective in predicting the effects of forgetting on task proficiency over no-practice intervals up to 1 year in duration (e.g., Hagman, Hayes, & Bierwirth, 1986; Rose et al., 1985; Rose, Radtke, Shettel, & Hagman, 1985). Since the accuracy of the predictions that the method provides depends directly on the reliability and validity of the rating instrument, the method is referred to here as the "Rating Instrument Method."

The second method depends on the use of task ratings and conjoint scaling methods (e.g., Krantz & Tversky, 1971; Nygren, 1982) to generate predictions about the effects of multidimensional variables like sleep loss. This method was used by the Workload and Ergonomics Branch of the Harry G. Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio, in the development of a technique for assessing mental workload--Subjective Workload Assessment Technique or SWAT (e.g., Reid & Nygren, 1988). This method is referred to here as the "Rating Scale Method." We feel that this method holds potential for application in this arena, but that it also may require a great deal more developmental effort than that required by the "Rating Instrument Method". Information on the Rating Scale Method is presented in Appendix B. The Rating Instrument Method is the basis for the recommended approach examined in the remainder of this report.

Rating Instrument Method

Applying the Rating Instrument Method to the sleep loss domain entailed performing the same basic steps that Rose et al., (1985) performed in their work on skill retention. These steps are described in the following paragraphs:

Identify critical dimensions for soldier performance. The first step entailed reviewing the scientific literature on sleep deprivation. The goal of this review was to identify those task characteristics known or suspected to influence performance in the absence of sleep. Three characteristics appeared especially critical to performance: mental effort load, time load, and motivation/arousal.

Mental effort load depends on the absolute amount of attentional capacity or effort required by the task and the duration of the task. This includes functions such as monitoring, retrieving information from memory, performing calculations, making decisions, and so on. Early experiments on the effects of sleep deprivation frequently yielded null results. Today it appears that these results were obtained largely because these experiments only imposed light, intermittent mental demands on subjects (e.g., Ainsworth & Bishop, 1971; Banks, Sternberg, Farrell, Debow, & Dalhamer, 1970). When mental demands were increased, either through increases in the attentional or thinking demands of the task, complexity of the task (e.g., uncertainty, unpredictability, unfamiliarity), or duration of work, clear evidence of the disruptive effects of sleep loss were obtained (e.g., Angus & Heslegrave, 1985; Babkoff, Thorne, Sing, Genser, Taube, & Hegge, 1985; Williams, Kearny & Lubin, 1965).

Time load refers to the amount of time available for an operator to perform a task. This includes both the overall time and rate at which the person must work to comply with task requirements. As an example, Williams and Lubin (1967) found that mental addition at a rate of one addition per 2 seconds did not show effects of two nights of sleep loss. However, mental addition was impaired after two nights of sleep loss when this rate was increased to one addition per 1.25 seconds.

The term "motivation/arousal" is used here to refer to characteristics of the task or task environment which influence a person's motivation to perform or ability to remain awake under conditions of sleep loss. For example, long monotonous tasks (e.g., monitoring) that lead to lowered arousal are among the most affected by sleep loss. On the other hand, variables that lead to states of heightened motivation/arousal, such as feedback or other incentives (e.g., Wilkinson, 1961), exercise (e.g., Englund, Ryman, Naitoh, Hodgdon, 1985), and noise (Wilkinson, 1963) are associated with improved performance under conditions of extended sleep loss.

Convert dimensions into questions (rating scales) with clearly defined answer options (anchors). Based on this review, an eight-question rating scale was developed (Appendix C). Questions 1, 2, and 5 were designed to measure mental effort load. Questions 3 and 4 were written to reflect time load. And, Questions 6, 7, and 8 were designed to measure a task's motivation/arousal characteristics. All but two of the items included five answer options.

Reference materials related to the questions included in the rating scale are noted in Appendix D. Some of references are to literature reviews that summarize supporting data (e.g., Belenky, Krueger, Balkin, Headley & Solick, 1987; Johnson, 1982; Krueger, 1989; Naitoh & Townsend, 1970). Other references are to original research which obtained results consistent with the use of specific questions or answer options. Some of this research was highlighted in the previous section.

Assign points to each question and each answer option. It was tentatively decided to assign equal weight to the three critical dimensions--mental effort load, time load, and motivation/arousal. Consequently a total of 100 points was assigned to Questions 1 and 2; 100 points was assigned to the combination of Questions 3 and 4; and 100 points was assigned to Questions 6, 7, and 8.

Generally speaking, the lower the demands in terms of mental effort or time that are imposed on the soldier, the more points were assigned to the answer option. Additionally, the more task characteristics appeared likely to raise motivation/arousal levels, the more points were assigned to the answer option. For example, tasks that can be performed more or less automatically, without conscious effort, are known to be largely impervious to the effects of sleep loss (e.g., Weiskotten & Ferguson, 1930). This is especially true of tasks which are wholly self-paced, that is, tasks where the subject controls the stimulus display (if any) and can respond at his leisure. These tasks may be performed more slowly, but they are much less likely to induce errors than work-paced tasks (e.g., Williams, Lubin, & Goodnow, 1959).

Some answer options were weighted far more heavily than others. For example, task monotony was seen as more important than the presence of feedback or some other incentive to performance under conditions of sleep loss. Similarly, time to task completion and rate considerations were seen as more critical than the number of break periods that occur throughout a test session. Some of these choices, such as task monotony, were suggested by various authors (e.g., Krueger, 1989). Others, such as time to task completion, were made because they seemed to provide the best fit to available data.

An additional plus or minus 25 points is possible (Question 5), depending on the answer to Question 4. If break periods do not occur regularly throughout a test session, a decrement in task performance can be anticipated (e.g., Angus & Heslegrave, 1985; Mullaney, Kripke, Fleck, & Johnson, 1983). This appears true even for tasks which are relatively short in duration, for example, one minute (e.g., Heslegrave & Angus, 1985). If break periods do occur regularly throughout a test session, no decrement in task performance usually is observed, particularly if the task is one of relatively short duration. However, a decrement in task performance is likely to be observed if the task must be performed for a relatively long duration (e.g., Wilkinson, 1961, 1964, 1968).

The process of developing questions and answer options and assigning points was treated analytically and iteratively. As the literature on sleep loss was reviewed, questions and answer options continued to evolve. So too did the points (or weights) assigned specific questions and answer options. Thus, some questions that were included originally were later dropped or given decreased emphasis. Others that were deemed less important originally were later added or given increased emphasis. As an example, the interest value of a task appears an important variable for sleep loss. In one experiment, a battle game was found so interesting that subjects were able to work at the game for an hour without showing the effects of over 50 hours of total sleep loss (Wilkinson, 1964). Initially, it was decided to include a question to assess the level of "interest" that a task generates, but then, later, this question was dropped because it was felt that, in the absence of a sound operational definition of the term "interest," this question may be the source of more error variance than predictive power. Overall, the goal was to develop a rating scale that would be easy to understand and simple to use (i.e., reliable) and that would provide an excellent fit to existing data (i.e., valid).

During the preliminary development and testing of the rating scale, special attention was given to research that included (1) clear descriptions of tasks and test conditions and (2) data on the effects of varying amounts of sleep loss on task performance (e.g., Angus & Heslegrave, 1985; Heslegrave & Angus, 1985; Thorne, Genser, Sing, &

Hegge, 1983; Weiskotten & Ferguson, 1930; Williams, Lubin, & Goodnow, 1959). Experimental tasks were rated based on the descriptions that were available. Once these ratings had been made, the results were subjectively evaluated against the available performance data. If the rating given a specific task appeared too high or too low relative to observed performance, the rating scale questions, answer options, or point structure was modified. Then all tasks were rated again using the revised rating scale. This process was repeated until the majority of the available data could be accounted for.

To carry out the process noted above, data that had been developed under a wide range of experimental conditions had to be combined. This required that performance scores be converted to a common metric. The metric that was employed was a "percent baseline" score. This approach was suggested by research performed by Underwood (1957). Using such an approach, Underwood (1957) was able to combine the results of some 14 separate studies to demonstrate a clear relationship between number of previous lists learned and amount of forgetting.

Baseline performance was defined either as the mean within-subject performance across the first 18 hours of sleep loss or control group performance (Weiskotten & Ferguson, 1930). Few performance decrements are observed across the first 18 hours without sleep. Then, performance usually falls in step-wise fashion with the onset of the early morning circadian cycle (e.g., Angus & Heslegrave, 1985; Belenky et al., 1987).

Establish rule for combining points. As tasks were being rated, points were combined additively. An additive model was followed because it is easy to understand and simple to use. We also had no reason for believing that a more complex combination rule would apply.

Establish function that relates combined score and time interval to soldier performance. Once the Sleep Loss Effects Task Rating Sheet had undergone preliminary development and testing, task ratings were developed. In all, 21 different tasks were rated. These ratings were accomplished after reviewing the research on sleep loss and finding reports of experiments on tasks ranging from the simple to the task performance after different lengths of time without sleep. And, in each case, these data were converted to percent baseline performance scores. The task ratings are presented in Appendix E along with these scores.

After the task ratings and percent baseline performance scores had been developed, they were submitted to a multiple regression analysis. The independent variables were Task Rating and Hours Without Sleep. Using the regression coefficients, we generated predicted values for the dependent variable (percent of baseline performance) for selected values of hours without sleep and task ratings. These predicted values are shown in Table 3. Other values may be computed using the regression equation generated by this analysis, as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 = 96.7163473 + (-0.95793196)X_1 + (0.20570088)X_2$$

where

Y = Predicted Performance Level as a Percent of Baseline Performance (Dependent Variable)

β_0 = Intercept (96.7163473)

X_1 = Hours without Sleep (Independent Variable)

X_2 = Task Rating (Independent Variable)

A similar type of procedure was used by Rose et al. (1985) in developing a system for predicting the effects of forgetting on the performance of different military tasks. A key difference, however, was that Rose et al. (1985) based their system on original data they collected. As noted earlier, our Sleep Loss Effects Prediction Matrix was developed using original task ratings and existing sleep loss performance data (Appendix E).

Given the Sleep Loss Effects Prediction Matrix, one need only carry out two steps to estimate the effects of sleep loss on any given task. The first step involves rating the task of interest using the Sleep Loss Effects Task Rating Sheet. It is possible for a task to be rated anywhere from -10 to 325, depending on its unique characteristics. The higher the rating score, the better performance is predicted to be under conditions of extended sleep loss. Thus, for example, we rated a cognitively demanding two-column addition task (Thorne et al., 1983) as 50. Performance on this task would be expected to suffer far more than performance on a simple ball tossing task, which we rated as 215 (Weiskotten & Ferguson, 1930) (see Appendix E).

The second step involves inserting a task's sleep loss effects task rating score into the Sleep Loss Effects Prediction Matrix (Table 3). The numbers along the left-hand column of the table are the sleep loss effects task rating scores; the numbers along the top row represent hours without sleep. The numbers in the body of the table represent percent baseline performance. Thus, given a rating score of 50, performance between 37 and 54 hours without sleep would be expected to equal approximately 55% of baseline performance.

Turning to Appendix E and reviewing data obtained by Thorne et al. (1983), it can be seen that actual performance during that time frame (44 to 48 hours without sleep) averaged approximately 60 percent of baseline performance. Similarly, given a rating of 215, performance at all intervals out to 72 hours would be expected to remain near baseline. Again, turning to Appendix E and considering data obtained by Weiskotten and Ferguson (1930), it can be seen that 72 hours of sleep loss had no apparent effect on the ball tossing task.

Table 3

Sleep Loss Effects Prediction Matrix

Total Score from Answer Sheet		Hours	Without	Sleep
	18	36	54	72
150	100	93	76	59
140	100	91	74	57
130	100	89	72	54
120	100	87	70	52
110	100	85	68	50
100	100	83	66	48
90	98	81	64	46
80	96	79	61	44
70	94	77	59	42
60	92	75	57	40
50	90	73	55	38
40	88	70	53	36
30	86	68	51	34

Assess inter-rater reliability and predictive power. During the initial development of the Sleep Loss Effects Task Rating Sheet, only one rater produced all the task ratings presented in Appendix E. As a result, no measure of inter-rater reliability was possible. A measure of inter-rater reliability is critical if more than one rater is expected to know how to use a rating instrument. For a rating instrument to demonstrate high inter-rater reliability, it must be simple to use and easy to understand. Instructions must be clear. Questions and answer alternatives must be unambiguous. And, any potentially confusing terms must be operationally defined. In short, the rating instrument must be designed in a way that limits the probability that individual raters will use it differently from one another.

Rose et al. (1985) reported high measures of inter-rater reliability for the rating instrument that they developed to predict task retention (e.g., $r = .90+$). They also demonstrated very positive results in tests of the predictive power of their rating instrument. For example, in one experiment, Rose et al. (1985) trained three groups of soldiers ($n = 140$) on 22 tasks to a criterion of one correct performance. Groups were then tested for retention either 2, 4, or 6 months later. A strong positive relationship was observed between actual and predicted performance at each retention test, with correlations being around 0.9 at the 2-month retention test and 0.7 at the other two retention tests. However, prior to having inexperienced raters try out their rating instrument, Rose et al. (1985) spent the time needed to prepare specific guidance on its use (e.g., Rose, Radtke, Shettel, & Hagman, 1985). We have proposed that such guidance be developed prior to a test of the Sleep Loss Effects Rating Sheet and that measures of inter-rater reliability and predictive power be taken later once this work is completed.

Results of the Multiple Linear Regression Analysis. At this point, only a preliminary attempt was made to measure the predictive power of the rating sheet. Data presented in Appendix E were submitted to a separate multiple regression analysis, where the independent variables were Task Rating and Hours Without Sleep.

The dependent variable was Percent Baseline Performance. Therefore, the multiple regression model used to analyze the data is as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$$

where:

Y	=	% Baseline (Dependent Variable)
β_0	=	Intercept
β_1	=	Hours without Sleep (Independent Variable)
β_2	=	Task Rating (Independent Variable)
e	=	Error

The statistical package used to generate the analysis is the Statview™ statistical package for the Macintosh computer.

Significance of Regression. In fitting the multiple linear regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_i,$$

Our test of the hypothesis,

$$H_0: \beta_1 = \beta_2 = 0$$

$$H_0: \beta_i \neq 0 \text{ for at least one } i$$

is depicted in Table 4.

Table 4

Analysis of Variance Table

Source of Variation	Sum of Squares	DF	Mean Square	F_0
Regression	67138.26	2	33569.13	196.24
Residual	29422.06	172	171.0585	
Total	96560.32	174		

Since $F_{0.01,2,172} = 4.61$, we reject H_0 and conclude that a least one β_i is significantly different from zero. The results of this analysis were positive, $\underline{F}(2, 172) = 196.24$, and the $r^2 = 0.6952$. (Appendix F).

This result is very encouraging. However, the predictive power of the Rating Sheet cannot help but be inflated to some degree. Key concerns are as follows:

- The Task Rating Sheet was developed using the same data that were used in the regression analysis. To obtain a more exacting measure the predictive power of the methodology, new task performance data are required. Further, these data should be collected under conditions where the experimenters are "blind" to the Task Rating Sheet predictions about the effects of sleep loss on the performance of specific tasks.

- All of the research that was used to support the analysis (Appendix E) was conducted under laboratory conditions.

- All of the tasks were rated by an individual with a complete knowledge of the intended meaning of the rating questions and answer options.

- Performance data in Appendix E represent group averages. Averaging data tends naturally to reduce error variability, which would enhance its predictive power.

A further concern is evidenced by the presence of Part II of the Sleep Loss Effects Task Rating Sheet--Relative Criticality Rating. There is a fair body of research which suggests that, under conditions of sleep loss and time pressure, performance may slow to a point where some tasks simply cannot be performed (e.g., Banderet, Stokes, Francesconi, Kowal, & Naitoh, 1981; Thorne et al., 1983: Williams & Lubin, 1967; Williams, Lubin, & Goodnow, 1959). For example, the Banderet et al. experiment involved having artillery fire direction center (FDC) teams participate in a sustained tactical battle operation. Team members worked on maps and plotted preplanned and unplanned targets, with concurrent fire missions that often were superimposed with calls for preplanned fire. Teams made more errors over time, but generally remained effective until they withdrew from the experiment. Significantly, however, self-initiated activities, such as revising pre-planned fire missions, were subject to rapid deterioration. Indeed, after 36 hours without sleep, many of these activities no longer were being performed at all.

Additional research is required to establish the types of tasks that are most likely to be left unperformed by sleep deprived soldiers. It also is important to know the conditions under which these tasks are likely to be left unperformed (e.g., hours without sleep). Part II of the Sleep Loss Effects Task Rating Sheet represents an initial attempt to identify these tasks. Given the results of the Banderet et al (1981) experiment, it is reasonable to hypothesize that key considerations would include the perceived criticality of the tasks and the extent to which task performance is seen as depending on personal initiative. As the Sleep Loss Effects Task Rating Sheet now stands, however, a task may not be judged susceptible to the effects of sleep loss (e.g., very low mental effort load) to a point when it simply stops being performed.

Step 5: Establish Means for Modifying Model Unit and Systems Effectiveness Variable Data Based on the Step 4 Predictions, Thereby Creating Performance Shaping Functions

The process of using the Step 4 predictions to modify model unit and systems effectiveness variable data involves four discrete operations. These operations are as follows:

1. Identify the unit and systems effectiveness variable data for the model that is to be modified.

2. Evaluate the task of interest using the Sleep Loss Effects Task Rating Sheet (Appendix C).

3. Determine the extent to which performance on the task of interest will be degraded using the Sleep Loss Effects Prediction Matrix.

4. Multiply the unit and systems effectiveness variable data by the predicted percent level of performance indicated in the Sleep Loss Effects Prediction Matrix.

The following is presented for illustration purposes. The example uses target acquisition as the task of interest.

Identify Unit and Systems Effectiveness Variable Data

Table 5 presents the VIC model target acquisition times by target status and range interval. The times were taken from the VIC Data Input and Methodology Manual, (1979). The target data are for a combat vehicle such as the T-72 main battle tank.

Table 5

Unit and Systems Effectiveness Variable Data -- Target Acquisition Times in Seconds -- VIC Model

Target Status	Range Interval (m)			
	1000	2000	3000	4000
Stationary				
Hull Defilade	3.4	6.9	10.4	30.6
Exposed	3.4	6.5	9.4	23.3
Moving	3.4	5.5	8.3	17.6

Evaluate the Task of Interest

The ability to generate reliable and valid task rating data depends heavily on the quality of the rating instrument. It also depends on the quantity and quality of the information that is available about the task to be rated. This is why subject matter experts usually are asked to perform task ratings. They are the people who are most familiar with the tasks of interest.

Producing reliable and valid task ratings also depends on having a clear understanding of the conditions under which the task is being performed and the standards to which it is being performed. A task may impose very different demands on a performer depending on how the task, conditions, and standards are defined. This is particularly true of a task like target acquisition, which can occur under a wide range of conditions and standards. Target acquisition generally is defined as the detection, identification, and location of a target in sufficient detail to permit the effective employment of weapons (Department of Defense, 1984). It is a continuing requirement for all tank crew members, whether in the offense or defense, moving or stationary. Some of the variables that have been cited as influencing target acquisition include scene (or total picture) variables (e.g., numbers, sizes, shapes, and distribution of areas contextually likely to contain the target object); target object variables (e.g., size, color, resolution; and observer variables (e.g., training) (Biberman, 1973).

For purposes of this example, subject matter experts were not used to evaluate the task of interest. A different set of ratings may have resulted if subject matter experts were used. Additionally, the target acquisition task was only rated under a single set of conditions to a single standard. Conditions and standards were defined very broadly. Conditions were defined simply as "very demanding; combat," and standards set simply at "maximum." Different ratings would be expected for the task if it were regarded as occurring under less demanding conditions or to a different standard.

The ratings that were produoed in response to the various questions on the Sleep Loss Effects Task Rating Sheet appear in Table 6. The task was seen as imposing high mental effort and time load demands. However, it was seen as being relatively unlikely to be disrupted by sleep loss, given a lack of monotony (target rich environment), the highest possible incentive for effective performance (survival), and a highlv stimulating task environment (combat).

Table 6

Sleep Loss Effects Task Rating Sheet Scores for Target Acquisition

Question		Score
1.	What are the mental or thinking requirements of the task?	0
2.	How complex is the task?	10
3.	How important are time or rate considerations to the successful performance of the task?	0
4.	How often do break periods of varying types occur throughout the test session?	0
5.	How long is the task performed without interruption?	0
6.	Is the task monotonous (the same response required to the same stimuli) or otherwise conducive to sleep?	20
7.	Is feedback or some other incentive used to motivate performers to try harder or persist longer at the task?	25
8.	Is the task environment conducive to sleep?	25
TOTAL		80

Use the Sleep Loss Effects Prediction Matrix

The Sleep Loss Effects Prediction Matrix was presented earlier (Table 3). Given the score of 80 that resulted from the task rating process, and assuming the performer has gone 36 hours without sleep, level of performance is expected to be at 79% baseline. At 54 hours without sleep, performance is expected to be at 61% baseline.

Dividing the unit and systems effectiveness variable data by the predicted fractional level of performance

Dividing the data in Table 5 by the predicted fractional level of performance (percent divided by 100) yields the predicted target acquisition times shown in Table 7. The assumptions here are that unit and systems effectiveness variable data included in the VIC model are representative of soldier baseline performance, and that the methods outlined above for generating performance shaping functions (i.e., percent baseline performance estimates) are valid. Both assumptions deserve more detailed consideration in the future.

Table 7

Predicted Target Acquisition Times (sec) by Range Interval and Target Status

Target Status	Range Interval (m)			
	1000	2000	3000	4000
Stationary				
Hull Defilade	4.3	8.7	13.2	38.7
Exposed	4.3	8.2	11.9	29.5
Moving	4.3	7.0	10.5	22.3

Only a portion of the unit and systems effectiveness data included in combat models are representative of soldier performance (e.g., Army Training and Evaluation Program results; field observations). Many models in the Army's inventory use weapons test data ("weapons baseline data") that are not relevant to consideration for man in-the-loop effects ("human performance baseline data"). The current approach is based on the assumption that data of interest are human performance baseline data.

Step 6: Recommend Possible Approaches for Adding the Performance Shaping Functions to a Combat Model

There are three alternative approaches for modifying combat model data to account for human performance:

1. Modify existing input data.

2. Create data look-up tables to account for the effects of select human performance variables and modify existing models to use them.

3. Develop new combat models which are designed from the outset to account for the effects of select human performance variables.

Alternative 1 may be the simplest and least expensive approach, but only for the short run. Using this approach, a great deal of effort would be required to account for the dynamic effects of soldier performance. To model change, new input data, commensurate with the status of the elements being modeled (the model state vector), would have to be input each time a new set of conditions was introduced. Simply capturing the effects of time on performance would require periodic model halts. On the surface, disadvantages associated with pursuing this alternative appear at least as great as the advantages.

Both Alternatives 2 and 3 would allow change to be modelled with considerably more elegance than Alternative 1. Presumably, both alternatives would enable this type of modelling to occur without the need for repeated intervention. In our opinion, however, Alternative 3 offers the likelihood of the greatest long-term returns on investments in the area. Current models are very complex. Efforts aimed at modifying these models are likely to be extensive and potentially far less cost effective than simply starting with a new design.

Alternatives 2 and 3 show more promise as long term solutions to modeling soldier performance than Alternative 1. However, early assessments of the impact of soldier performance variables on combat model outputs may be obtained by testing Alternative 1.

RESULTS AND DISCUSSION

The purpose of this research was to develop a method for including soldier performance considerations in Army combat models. This goal has been achieved with a positive result. The following six-step process is proposed:

Step 1: Identify Candidate Combat Model and Model Processes

The VIC combat model was selected to illustrate the proposed methodology. The model is representative of the Army's combat model inventory and includes variables potentially amenable to the effects of human performance. Also, the data are specific and highly detailed.

Step 2: Identify a Set of Model Unit and Systems Effectiveness Variables that can be Influenced by Soldier Performance

This process was accomplished by separating the variables that can be influenced by soldier performance from those that are strictly scenario driven (e.g., number of red and blue weapon systems) or that are determined by the engineering characteristics of the weapon systems being modelled (e.g., Target vulnerability).

Step 3: Select a Candidate Soldier Performance Variable

Sleep loss was chosen as a candidate variable from among the many possible soldier performance variables. Sleep loss was selected because it is known to be a potent variable for human performance. Additionally, a fair amount of research has been performed attempting to account for the effects of sleep loss on human performance. The availability of this research greatly aided the development of the Sleep Loss Effects Rating Sheet.

Another reason for focusing initially on sleep deprivation instead of a wider range of performance variables was to keep our approach as simple as possible. This, we believed, was necessary given the complexity of the problem we were facing and the hope of one day extending the approach to actual application. Yet, it quickly became evident, even in dealing only with sleep deprivation, that "keeping it simple" would be a significant challenge. Sleep loss effects are not a simple product of a single variable, such as hours without sleep. They appear the result of a host of different variables acting alone and in combination with one another. And, they cannot be predicted in the absence of a method that can deal with complexity.

How much complexity the method is capable of dealing with is another issue. Obviously, in working to increase the fidelity of our combat models, we must account for more variables than those associated with sleep deprivation. But how many variables can we handle with any degree of precision? And, how can the method be expanded to account for these variables?

At this point, we only can suggest possible answers to these questions. First, the number of variables that can be dealt with will depend on how one defines the word "variable". For example, "sleep deprivation" may be regarded as a single variable, or it may be regarded as a composite of many variables. Our belief is that the variables of interest to people in the combat modeling community will tend to be composites (e.g., stress) and that we would be doing well to deal effectively with two or three such "variables". The number of variables that can be dealt with also will depend on the amount of data that are available both to guide the development of an expanded rating instrument and to permit some amount of preliminary testing. As noted earlier, the rating instrument development process is very much an iterative, trial-and-error process. The more data that are available to direct this process, the faster and better it will be. Ultimately, the answer to the first question will depend on the number of variables that it is cost-effective to deal with. If with a very small number of very salient, well-researched variables, it may not make sense to try for more.

One of the best features of the Rating Instrument Method is its ability to be expanded as needs dictate. It is not simply a matter of adding more questions. Weights assigned to questions and response alternatives also have to be adjusted. However, there is no fixed limit on the number of questions that can be included. The problem is not in expanding the rating instrument; the problem is knowing how to expand the rating instrument so that the results which are produced are reliable and valid.

Step 4: Define a Method for Predicting How Soldier Performance will be Affected by the Candidate Variable

Two alternative methods were proposed for generating specific predictions about the effects of particular variables on soldier performance. Both methods depend on the use of task ratings. One method was developed originally to support predictions about the effects of forgetting on task performance (Rose et al., 1985). It was referred to here as the Rating Instrument Method. The other method was developed originally to aid in estimating the mental workload associated with performing specific tasks (e.g., Reid & Nygren, 1988). It was referred to as the Rating Scale Method. The Rating Instrument Method was proposed as the method of first choice, primarily because it appears better suited to handling the range of variables that must be considered to effectively model soldier performance. Both methods are well-grounded in research, and both may be regarded as viable candidates until proven otherwise.

Step 5: Establish Means for Modifying Model Unit and Systems Effectiveness Variable Data Based on the Step 4 Predictions, Thereby, Creating Performance Shaping Functions

The process of using the Step 4 predictions to modify model unit and systems effectiveness variable data was seen as involving four discrete operations. These operations were as follows:

1. Identify the unit and systems effectiveness variable data for the model that is to be modified.

2. Evaluate the task of interest using the Sleep Loss Effects Task Rating Sheet (Appendix C).

3. Determine the extent to which performance on the task of interest will be degraded using the Sleep Loss Effects Prediction Matrix or the Regression Equation.

4. Multiply the unit and systems effectiveness variable data by the predicted percent level of performance indicated in the Sleep Loss Effects Prediction Matrix.

Step 6: Recommend Possible Approaches for Adding the Performance Shaping Functions to a Combat Model

Three alternative approaches were suggested for modifying combat model data to account for human performance:

1. Modify existing input data.

2. Create data look-up tables to account for the effects of selected human performance variables and modify existing models to use them.

3. Develop new combat models which are designed from the outset to account for the effects of selected human performance variables.

The first alternative was regarded as a possible short-term solution. However, the second and third alternatives were seen as far more effective in capturing the dynamic nature of human performance. Overall, Alternative 3 was viewed as providing the best long-term returns on investments in the area.

ISSUES AND DIRECTIONS FOR FUTURE RESEARCH

The need to account for soldier performance effects in combat models is controversial. Some attempts have been made to validate this need by conducting runs of different types on existing models (e.g., CASTFOREM). A possible problem with this approach is that it depends on the use of existing combat models and the data that are resident in these models. It is not clear that an exacting test of the effects of soldier performance variables can be conducted in this manner.

Future research must place more emphasis on the selection of a model for demonstration testing and the choice of data that are used to model soldier performance. This may entail establishing criteria for selecting one combat model over another and for one modeling scenario over another. It also should entail paying special attention to the human performance variables that are selected for modeling, the levels at which these variables are set, and the validity of model input data.

There is a wide variety of human performance variables that must be regarded as candidates for future modeling work. Some means for prioritizing these variables is required. One very pragmatic approach is to establish these priorities on the basis of available data. If sufficient data are available to allow accurate predictions about the effects of a specific variable, and if the variable appears well linked to combat performance, the variable would be regarded as a good candidate for modeling. Otherwise, it probably would not, at least not at this time. Another potential means for prioritizing soldier performance variables may be to interview soldiers returning from combat in the Persian Gulf. These interviews could be used to establish the relative importance of particular variables and help give direction to future work in the area.

What happens in cases where a variable is seen as being a key determinant of combat performance but relatively little empirical data are available to support the development of valid rating instrument (e.g., Sleep Loss Effects Task Rating Sheet)? As one example, relatively little research is available about the effects of less than full sleep loss. Yet, the average soldier expects to receive as least some sleep each night (e.g., Van Nostrand, 1988). Can a valid instrument be developed anyway?

The answer to this question is probably yes, although the development process is bound to be more difficult and subject to more criticism where data are lacking than where they are more plentiful. For this reason, we are inclined to focus initially on variables which have been well researched. A rating instrument that is based in research is much easier to defend than one that is based strictly on opinion

REFERENCES

Ainsworth, L. L., & Bishop, H. P. (1971). The effects of a 48-hour period of sustained field activity on tank crew performance (HumRRO Tech. Rep. No. 71-16). Alexandria, VA: Human Resources Research Organization.

Angus, R. G., & Heslegrave, R. J. (1985). Effects of sleep loss on sustained cognitive performance during a command and control simulation. Behavior Research Methods, Instruments, & Computers, 17, 55-67.

Babkoff, H.., Thorne, D. R., Sing, H. C., Genser, S. G., Taube, S. L., & Hegge, F. W. (1985). Dynamic changes in work/rest duty cycles in a study of sleep deprivation. Behavior Research Methods, Instruments, and Computers, 17, 604-613.

Banderet, L. E., Stokes, J. W., Francesconi, R., Kowal, D. M., & Naitoh, P. (1981). Artillery teams in simulated sustained combat: Performance and other measures. In L. C. Johnson, D. I. Tepas, W. F. Colquhoun, & M. J. Colligan (Eds.), The twenty-four hour workday: Proceedings of a symposium on variations in work-sleep schedules (DHHS NIOSH Pub. No. 81-127). Cincinnati, OH: U.S. Department of Health and Human Services, National Institute for Occupational Safety and Health.

Banks, J. H., Sternberg, J. J., Farrell, J. P., Debrow, C. H., & Dalhamer, W. A. (1970). Effects of continuous military operations on selected tasks (BESRL) Tech. Rep. No. 1166). Arlington, VA: U.S. Army Behavioral and Systems Research Laboratory.

Belenky, G. L., Krueger, G. P., Balkin, T. J., Headly, D. B., & Solick, R. E. (1987, April). Effects of continuous operations (CONOPS) on soldier and unit performance: Review of the literature and strategies for sustaining the soldier in CONOPS (WRAIR-BB-87-1). Washington, DC: Walter Reed Army Institute of Research.

Biberman, L. M. (1973). Image quality. In L. M. Biberman (ed.), Perception of displayed information. New York: Plenum.

Crew System Ergonomics Information Analysis Center. (1990, Fall). Subjective Workload Assessment Technique (SWAT). Wright-Patterson Air Force Base, Ohio.

Department of Defense. (1984). Dictionary of Military and Associated Terms (JCS Pub 1). Washington, DC: Joint Chiefs of Staff.

Department of the Army (1979a). VIC Executive Summary. Washington DC: Author.

Department of the Army (1979b). VIC Data Input and Methodology Manual. Washington DC: Author.

Englund, C. E., Ryman, D. H., Naitoh, P., & Hodgdon, J. A. (1985). Cognitive performance during successive sustained physical work episodes. Behavior Research Methods, Instruments, & Computers, 17, 75-85.

Donnell, J. M. (1969). Performance decrement as a function of total sleep loss and task duration. Perceptual & Motor Skills, 29, 711-714.

Hagman, J. D., Hayes, J. F., & Bierwirth, W. (1986). A method for estimating task retention: Research pays off. Army Trainer, 6, 54-55.

Heslegrave, R. J., & Angus, R. G. (1985). The effects of task duration and work-session location on performance degradation induced by sleep loss and sustained cognitive work. Behavior Research Methods, Instruments, & Computers, 17, 592-603.

Johnson, L. C. (1982). Sleep deprivation and performance. In W. B. Webb (Ed.) Biological Rhythms, Sleep, and Performance. New York: Wiley.

Johnson, R. M. (1973). Pairwise nonmetric multidimensional scaling. Psychometrika, 38, 11-18.

Krantz, D. H., & Tversky, A. (1971) Conjoint measurement analysis of composition rules in psychology. Psychological Review, 78, 151-169.

Krueger, G. P. (1989). Sustained work, fatigue, sleep loss and performance: A review of the issues. Work & Stress, 3, 129-141.

Kruskal, J.B. (1965). Analysis of factorial experiments by estimating monotone transformations of data. Journal of the Royal Statistical Society, 27, 251-263.

McNally, R. E., Machovec, A. M., Ellsy, D. T., & Hursh, S. R. (1989). Evaluation of sleep discipline in sustaining unit performance (Contract No. MDA903-88-D-1000). McLean, VA: Science Applications International Corporation.

Miller and Bonder, S. (1982). Human factors representations for combat models (ARI TR 571).

Mullaney, D. J., Kripke, D. F., Fleck, P. A., & Johnson, L. C. (1983). Sleep loss and nap effects on sustained continuous performance. Psychophysiology, 20, 643-651.

Naitoh, P., & Townsend, R. E. (1970). The role of sleep deprivation research in human factors. Human Factors, 12, 525-585.

Nygren, T. E. (1982, April). Conjoint measurement and conjoint scaling: A user's guide (AFAMRL-TR-82-22). Air Force Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. Human Mental Workload. North Holland: Elsevier Science.

Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock & N. Meshkati (Eds.) Human mental workload. New York: North Holland.

Reid, G. B., Shingledecker, C. A., & Eggemeier, F. T. (1981). Application of conjoint measurement to workload scale development. Proceedings of the Human Factors Society 25th Annual Meeting, Rochester, NY, October, 12-16, 1981.

Rose, A. M., Czarnolewski, M. Y., Gragg, F. E., Austin, S. H., Ford, P., Doyle, J., & Hagman, J. D. (1985). Acquisition and retention of soldiering skills (Tech. Rep. 671). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Rose, A. M., Radtke, P. H., Shettel, H. H., & Hagman, J. D. (1985). User's manual for predicting military task retention (Res. Prod. 85-26). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Thorne, D., Genser, S., Sing, H., & Hegge, F. (1983, May 2-4). Plumbing human performance limits during 72 hours of high task load. In Defense Group Proceedings: The human as a limiting element in military systems (Vol. 1, DS/A/DR[83]170). Toronto: NATO Defence Research Group.

Underwood, B. J. (1957). Interference and forgetting. Psychological Review, 64, 49-60.

Vandivier, P. L. (1990). Soldier dimensions in combat models (ACN 86098). Fort Benjamin Harrison, In: TRADOC Analysis Command.

Van Nostrand, S. J. (1986). Model effectiveness as a function of personnel [ME = f(PER)] (Research Paper CAA-SR-86-34). Bethesda, MD: U.S. Army Concepts Analysis Agency.

Van Nostrand, S. J. (1988). Including the soldier in combat models (Executive Research Project S73) Fort McNair, DC: The Industrial College of the Armed Forces.

Weiskotten, T. F., & Ferguson, J. E. (1930). A further study of the effects of sleep. Journal of Experimental Psychology, 13, 247-266.

Wilkinson, R. T. (1969). Some factors influencing the effect of environmental stressors upon performance. Psychological Bulletin, 72, 260-272.

Wilkinson, R. T. (1968). Sleep deprivation: Performance tests for partial and selective sleep deprivation. In L. A. Abt & B. F. Reiss (Eds.), Progress in clinical psychology. New York: Grune & Statton.

Wilkinson, R. T. (1965). Sleep deprivation. In O. G. Edholm & A. L. Bacharach (Eds.), The physiology of human survival. New York: Academic.

Wilkinson, R. T. (1964). Effects of up to 60 hours of sleep deprivation on different types of work. Ergonomics, 17, 175-186.

Wilkinson, R. T. (1963). Interaction of noise with knowledge of results, repeated testing, and individual differences. Journal of Experimental Psychology, 66, 332-337.

Wilkinson, R. T. (1961). Interaction of lack of sleep with knowledge of results, repeated testing, and individual differences. Journal of Experimental Psychology, 62, 263-271.

Williams, H. L., Kearney, O. F., & Lubin, A. (1965). Signal uncertainty and sleep loss. Journal of Experimental Psychoiogy, 69, 401-407.

Williams, H. L., & Lubin, A. (1967). Speeded addition and sleep loss. Journal of Experimental Psychology, 73, 313-317.

Williams, H. L., Lubin, A., & Goodnow, J. J. (1959). Impaired performance with acute sleep loss. Psychological Monographs: General & Applied, 73, 1-26.

APPENDIX A

VIC SYSTEMS EFFECTIVENESS VARIABLES POSSIBLY INFLUENCED BY SOLDIER PERFORMANCE - MANEUVER UNIT COMBAT

The manuever unit combat model process for the VIC model yielded ninety-three (93) systems effectiveness variables based on a preliminary assessment of the data input variables described in the VIC Data Input and Methodology Manual. Further assessment, based upon separating those variables that can be influenced by soldier performance from those that are driven strictly by the scenario (e.g., number of red and blue weapon systems) or determined by the engineering characteristics of the weapon systems in use (e.g., target vulnerability), identified sixteen (16) that could possibly be influenced by soldier performance as follows:

- Tactical Weapon Speed
- Acquisition Rate Factor for Moving Firer
- Minimum Threshold for Direct Fire Suppression
- Level Indirect Fire Suppression
- Probability of Acquisition in Infinite Time
- Proportion of Fire Vs False Targets
- Fraction Time Firing
- Kill Rates for Firer
- Fire Rate Factor
- Delay in Switching from Wide to Narrow Field of View
- Maximum Threshold for Direct Fire Suppression
- Factor for Visual Acquisition Rate for Blue/Red
- Mean Acquisition Time in Single Field of View
- Weapon Firing Rates
- Search Cutoff Time
- Fraction Time Moving

APPENDIX B

RATING SCALE METHOD

During the design, development, and test and evaluation of any advanced aircraft, the capabilities and limitations of the aircrew must be considered. Care must be taken that the new system does not place unreasonable demands on crew members by overwhelming them with too much information and too little time to process that information. Such considerations are often characterized by assessments of mental workload.

The Subjective Workload Assessment Technique (SWAT) was designed to measure mental workload, so the Rating Scale Method is most easily described in terms of this construct. This methodology holds potential for application to any domain, such as sleep loss, which is multidimensional in nature. The method depends on a two-step procedure: (1) scale development and (2) event scoring.

<u>**Scale development**</u>. Mental workload is proposed to be explained by three component factors: mental effort load, time load, and psychological stress load. Each of these factors is addressed at three different levels. Definitions for the three levels of each factor are as follows (Reid, Shingledecker, & Eggemeier, 1981, p. 523):

- **Mental Effort Load**

  1. Little conscious mental effort or planning required. Low task complexity such that tasks are often performed automatically.

  2. Considerable conscious mental effort or planning required. Moderately high task complexity due to uncertainty, unpredictability, or unfamiliarity.

  3. Extensive mental effort and skilled planning required. Very complex tasks demanding total attention.

- **Time Load**

  1. No or very few interruptions in the planning, execution, or monitoring of tasks. Spare time exists between many tasks.

  2. Task planning, execution and monitoring are often interrupted. Little spare time. Tasks occasionally occur simultaneously.

  3. Task planning, execution and monitoring are interrupted most of the time. No spare time. Tasks frequently occur simultaneously. Considerable difficulty in accomplishing all tasks.

- **Psychological Stress Load**

  1. Little risk, confusion, frustration, or anxiety exists and can be easily accommodated.

  2. The degree of risk, confusion, frustration, or anxiety noticeably adds to workload and requires significant compensation to maintain adequate performance.

  3. The level of risk, confusion, frustration, or anxiety greatly increases work load and requires tasks to be performed only with the highest level of determination and self-control.

Given the above, the mental workload represented by any particular hypothetical activity is defined in terms of a specific combination of the three levels of each factor (i.e., mental effort load, time load, and psychological stress load). In total, there are 27 such combinations, and the first step in the scale development procedure is simply to have subjects rank order the 27 combinations according to their perceived workload. For example, "1-1-1" and "3-3-3" would be at opposite ends of the continuum from each other, with "1-1-2", "-2-2-3", "3-3-2", and so on falling somewhere between these points. The results of this ranking then are transformed into an interval scale of workload ranging from 0 to 100. This transformation is accomplished by means of a psychometric technique known as numerical conjoint scaling (Krantz & Tversky, 1971; Nygren, 1982). Conjoint scaling techniques are designed to assess the joint effects of several factors and to extract the rule or composition principle that relates the factors to one another. A major advantage of the approach is that only ordinal data are required to produce an interval level scale which represents the joint effects of the factors.

The scaling routine in SWAT that is used to establish an interval scale of mental workload is based on modifications of two nonmetric scaling algorithms, MANANOVA (Kruskal, 1965) and NONMETRG (Johnson, 1973). Nonmetric scaling methods differ from metric scaling procedures in that they do not assume a linear relationship between observed data and final scale values. With nonmetric procedures, one does not need to assume that the respondent can and will make reliable ratings that have interval-scale properties when judging a complex construct like mental workload. A nonmetric scaling procedure only requires the data to be reliably rank ordered. A detailed description of the manner in which these scaling algorithms are used in SWAT is beyond the scope of this report. However, it is the subject of a recent book chapter entitled, "The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload" (Reid & Nygren, 1988).

Event scoring. During the event scoring step, tasks are rated using the same descriptors as were used for scale development. Thus, a pilot might be asked to rate a task such as a landing by assigning a 1, 2, or 3 to mental effort load, time load, and psychological stress load. Once this rating has been made, the 0-to-100 scale value corresponding to this rating is assigned as the workload value for that activity.

For purposes of the present work, the Rating Scale Method is seen as a possible alternative to the Rating Instrument Method for developing task ratings. The Rating Scale Method does not eliminate the need to develop task ratings or to relate those ratings to actual performance. However, the method is potentially more defensible from a purely psychometric standpoint than the Rating Instrument method. If the Rating Scale Method has a drawback, it is in its relative difficulty of use. Yet, even this drawback is potentially of limited consequence, given advances in efforts to automate the scale development process (See Crew System Ergonomics Information Analysis Center Gateway, 1990).

Applying the Rating Scale Method to the sleep loss domain first would entail identifying the dimensions underlying sleep loss. Earlier, it was observed that the dimensions mental effort load, time load, and motivation/arousal provide a reasonable fit to the data. The next step then would entail developing definitions for the various levels of these dimensions. Once these steps have been completed, use of the method would involve following normal scale development and event scoring procedures. Of course, as suggested above, task scores resulting from these procedures also would have to related to actual performance scores in order to generate specific performance predictions.

APPENDIX C

SLEEP LOSS EFFECTS TASK RATING SHEET

Part I: Task Performance Rating

Purpose

The purpose of Part I of this rating sheet is to aid predictions about the susceptibility of specific tasks to the effects of sleep loss.

1. **What are the mental or thinking requirements of the task? (50)**

   0 Very large; demands total attention (e.g., vigilance); full cognitive work load in terms of thinking, planning, problem solving, memorizing, etc. (e.g., logical reasoning)

   10 Large

   20 Moderate

   30 Small

   50 Very small; task may be performed automatically (e.g., road march)

2. **How complex is the task (50)**

   0 Very high task complexity (e.g., great uncertainty, unpredictability, unfamiliarity) (e.g., logical reasoning)

   10 High task complexity

   20 Moderate task complexity

30 Low task complexity

50 Very low task complexity (e.g., great certainty, predictability, familiarity) (e.g., signing one's name)

3. **How important are <u>time</u> or <u>rate</u> considerations to the successful performance of this task? (75)**

0 Of very great importance; heavy time pressure; work output can never be allowed to vary without risk of penalty (e.g., vigilance)

10 Of great importance; work output can be varied to a small degree without risk of penalty

25 Of moderate importance; work output can be varied to a moderate degree without risk of penalty

50 Of little importance; work output can be varied to a large degree without risk of penalty

75 Not important; performer can respond more or less at his (or her) leisure

4. **How often do break periods of varying types occur throughout the test session? (25)**

0 Very infrequently; intense work load conditions (Skip Question 5)

5 Infrequently (Skip Question 5)

10 Moderately often (Skip Question 5)

15 Frequently (Answer Question 5)

25 Very frequently; work paced to allow for substantial periods of rest (without sleep) (Answer Question 5)

5. **How long is the task performed without interruption? (25)**

Subtract 25 Relatively long duration (e.g., 30 minutes or longer)

0 Moderate duration

Add 25 Relatively short duration (e.g., 2 minutes or less)

6. **Is the task monotonous (the same response required to the same stimuli) or otherwise conducive to sleep?**

0 To a very large extent (e.g., highly repetive, never ending, boring)

10 To a large extent

20 To a moderate extent

30 To a small extent

50 To a very small extent (e.g., fun, interesting, stimulating)

7. **Is feedback or some other incentive used to motivate performers to try harder or persist longer at the task? (25)**

0 No

25 Yes

8. **Is the task environment conducive to sleep? (25)**

0 To a very large extent (e.g., safe, quiet, comfortable)

5 To a large extent

10 To a moderate extent

15 To a small extent

25 To a very small extent (e.g., unsafe, noisy, uncomfortable)

Part II: Relative Criticality Rating (Optional)

Purpose

The purpose of Part II of this rating sheet is to identify tasks likely to be left unperformed in the presence of increasing amounts of sleep loss and time pressure.

9. **Relative to other tasks (critical or otherwise), how important is this task?**

1 Of very little importance; this task probably would be among the first to be dropped in the presence of increasing amounts of sleep loss and time pressure

2 Of little importance

3 Of moderate importance

4 Of great importance

5 Of very great importance; this task probably would be among the last to be dropped in the presence of increasing amounts of sleep loss and time pressure

10. **Relative to other tasks, how much does the performance of this task depend strictly on personal initiative?**

1 To a very large extent

2 To a large extent

3 To a moderate extent

4 To a small extent

5 To a very small extent

11. **Relative to other tasks, how much is the non-performance of this task likely to: (a) jeopardize human safety, (b) threaten mission outcome, or (c) cause costly equipment malfunctions or delays?**

1 To a very small extent

2 To a small extent

3 To a moderate extent

4 To a large extent

5 To a very large extent

APPENDIX D

SLEEP LOSS EFFECTS TASK RATING SHEET REFERENCES

Part 1

1. **What are the mental or thinking requirements of the task?**

Babkoff, H., Thorne, D. R., Sing, H. C., Genser, S. G., Taube, S. L., & Hegge, F. W. (1985). Dynamic changes in work/rest duty cycles in a study of sleep deprivation. Behavior Research Methods, Instruments, and Computers, 17, 604-613.

Belenky, G. L., Krueger, G. P., Balkin, T. J., Headly, D. B., & Solick, R. E. (1987, April). Effects of continuous operations (CONOPS) on soldier and unit performance: Review of the literature and strategies for sustaining the soldier in CONOPS (WRAIR-BB-87-1). Washington, DC: Walter Reed Army Institute of Research.

Haslam, D. R. (1981). The military performance of soldiers in continuous operations: Exercises 'Early Call' I and II. In L. C. Johnson, D. I. Tepas, W. P. Colquhoun, & M. J. Colligan (Eds.), Biological Rhythms, Sleep and Shift Work. New York: Spectrum.

Haslam, D. R. (1985). Sustained operations and military performance. Behavior Research Methods, Instruments, and Computers, 17, 90-95.

Haslam, D. R., & Abraham, P. (1987). Sleep loss and military performance. In G. L. Belenky (Ed.), Contemporary Studies in Combat Psychiatry. New York: Greenwood.

Johnson, L. C. (1982). Sleep deprivation and performance. In W. B. Webb (Ed.) Biological Rhythms, Sleep and Performance. New York: Wiley.

Kleitman, N. (1939). Sleep and wakefulness. Chicago, IL: University of Chicago.

Kopstein, F., Siegel, A., Conn, J., Caviness, J., Slifer, W., Ozkaptan, H., & Dyer, F. (1985). Soldier performance in continuous operations (ARI Res. Note 85-68). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Krueger, G. P. (1989). Sustained work, fatigue, sleep loss and performance: A review of the issues. Work & Stress, 3, 129-141.

Laslett, H. R. (1924). An experiment on the effects of the loss of sleep. Journal of Experimental Psychology, 7, 45-58.

Laslett, H. R. (1928). Experiments on the effects of the loss of sleep. Journal of Experimental Psychology, 11, 370-396.

Naitoh, P., & Townsend, R. E. (1970). The role of sleep deprivation research in human factors. Human Factors, 12, 525-585.

Thorne, D., Genser, S., Sing, H., & Hegge, F. (1983, May 2-4). Plumbing human performance limits during 72 hours of high task load. In Defense Group Proceedings: The human as a limiting element in military systems (Vol. 1, DS/A/DR[83]170). Toronto: NATO Defense Research Group.

Wilkinson, R. T. (1964). Effects of up to 60 h urs of sleep deprivation on different types of work. Ergonomics, 17, 175-186.

Williams, H. L., Gieseking, C. F., & Lubin, A. (1966). Some effects of sleep loss on memory. Perceptual and Motor Skills, 23, 1287-1293.

Williams, H. L., & Williams, C. L. (1966). Nocturnal EEG profiles and performance. Psychophysiology, 3, 164-175.

2. **How complex is the task?**

Babkoff, H., Mikulincer, M., Caspy, T., Kempinski, D., & Sing, H. (1988). The topology of performance curves during 72 hours of sleep loss: A memory and search task. The Quarterly Journal of Experimental Psychology, 40, 737-756.

Herman, E. S., & Robinson, S. O. (1922). Effects of the loss of sleep. Journal of Experimental Psychology, 5, 19-22.

Naitoh, P., & Townsend, R. E. (1970). The role of sleep deprivation research in human factors. Human Factors, 12, 525-585.

Robinson, E. S., & Robinson, F. (1922). Effects of sleep loss. II. Journal of Experimental Psychology, 5, 93-100.

Williams, H. L., Kearney, O. F., & Lubin, A. (1965). Signal uncertainty and sleep loss. Journal of Experimental Psychology, 69, 401-407.

Williams, H. L., & Lubin, A. (1967). Speeded addition and sleep loss. Journal of Experimental Psychology, 73, 313-317.

Williams, H. L., Lubin, A., & Goodnow, J. J. (1959). Impaired performance with acute sleep loss. Psychological Monographs: General & Applied, 73, 1-26.

Wilkinson, R. T. (1964). Effects of up to 60 hours of sleep deprivation on different types of work. Ergonomics, 17, 175-186.

Wilkinson, R. T. (1965). Sleep deprivation. In O. G. Edholm & A. L. Bacharach (Eds.), The physiology of human survival. New York: Academic.

3. **How important are time or rate considerations to the successful performance of the task?**

Naitoh, P., & Townsend, R. E. (1970). The role of sleep deprivation research in human factors. Human Factors, 12, 525-585.

Patrick, G. T. W., & Gilbert, J. A. (1896). On the effects of loss of sleep. Psychological Review, 3, 469-483.

Schein, E. H. (1957). The effects of sleep deprivation on performance in a simulated communication task. Journal of Applied Psychology, 41, 247-252.

Weiskotten, T. F., & Ferguson, J. E. (1930). A further study of the effects of sleep. Journal of Applied Psychology, 41, 247-252.

Wilkinson, R. T. (1961). Interaction of lack of sleep with knowledge of results, repeated testing, and individual differences. Journal of Experimental Psychology, 62, 263-271.

Williams, H. L., & Lubin, A. (1967). Speeded addition and sleep loss. Journal of Experimental Psychology, 73, 313-317.

Williams, H. L., Lubin, A., & Goodnow, J. J. (1959). Impaired performance with acute sleep loss. Psychological Monographs: General & Applied, 73, 1-26.

4. **How often do break periods of varying types occur throughout the test session?**

Angus, R. G., & Heslegrave, R. J. (1985). Effects of sleep loss on sustained cognitive performance during a command and control simulation. Behavior Research Methods, Instruments, & Computers, 17, 55-67.

Heslegrave, R. J., & Angus, R. G. (1985). The effects of task duration and work-session location on performance degradation induced by sleep loss and sustained cognitive work. Behavior Research Methods, Instruments, & Computers, 17, 592-603.

Mullaney, D. J., Fleck, P. A., Okudiaira, N., & Kripke, D. F. (1985). An automated system for administering continuous workload and for measuring sustained continuous performance. Behavior Research Methods, Instruments, & Computers, 17, 16-18.

Mullaney, D. J., Kripke, D. F., Fleck, P. A., & Johnson, L. C. (1983). Sleep loss and nap effects on sustained continuous performance. Psychophysiology, 20, 643-651.

Mullaney, D. J., Kripke, D. F., Fleck, P. A. & Okudaira, N. (1983). Effects of sustained continuous performance on subjects working alone and in pairs. Perceptual and Motor Skills, 57, 819-832.

Naitoh, P., & Townsend, R. E. (1970). The role of sleep deprivation research in human factors. Human Factors, 12, 525-585.

5. How long is the task performed without interruption?

Ainsworth, L. L., & Bishop, H. P. (1971). The effects of a 48-hour period of sustained field activity on tank crew performance (HumRRO Tech. Rep. No. 71-16). Alexandria, VA: Human Resources Research Organization.

Babkoff, H., Thorne, D. R., Sing, H. C., Genser, S. G., Taube, S. L., & Hegge, F. W. (1985). Dynamic changes in work/rest duty cycles in a study of sleep deprivation. Behavior Research Methods, Instruments, and Computers, 17, 604-613.

Donnell, J. M. (1969). Performance decrement as a function of total sleep loss and task duration. Perceptual & Motor Skills, 29, 711-714.

Edwards, A. S. (1941). Effects of the loss of 100 hours of sleep. American Journal of Psychology, 54, 80-91.

Kleitman, N. (1923). On the effects of the loss of sleep. American Journal of Physiology, 67, 67-92.

Laslett, H. R. (1928). Experiments on the effects of the loss of sleep. Journal of Experimental Psychology, 11, 370-396.

Heslegrave, R. J., & Angus, R. G. (1985). The effects of task duration and work-session location on performance degradation induced by sleep loss and sustained cognitive work. Behavior Research Methods, Instruments, & Computers, 17, 592-603.

Thorne, D., Genser, S., Sing, H., & Hegge, F. (1983, May 2-4). Plumbing human performance limits during 72 hours of high task load. In Defense Group Proceedings: The human as a limiting element in military systems (Vol. 1, DS/A/DR[83]170). Toronto: NATO Defense Research Group.

Warren, N., & Clark, B. (1937). Blocking in mental and motor tasks during a 65-hour vigil. Journal of Experimental Psychology, 21, 97-105.

Williams, H. L., Kearney, O. F., & Lubin, A. (1965). Signal uncertainty and sleep loss. Journal of Experimental Psychology, 69, 401-407.

Williams, H. L., Lubin, A., & Goodnow, J. J. (1959). Impaired performance with acute sleep loss. Psychological Monographs: General & Applied, 73, 1-26.

Wilkinson, R. T. (1961). Interaction of lack of sleep with knowledge of results, repeated testing, and individual differences. Journal of Experimental Psychology, 62, 263-271.

Wilkinson, R. T. (1964). Effects of up to 60 hours of sleep deprivation on different types of work. Ergonomics, 17, 175-186.

Wilkinson, R. T. (1965). Sleep deprivation. In O. G. Edholm & A. L. Bacharach (Eds.), The physiology of human survival. New York: Academic.

Wilkinson, R. T. (1968). Sleep deprivation: Performance tests for partial and selective sleep deprivation. In L. A. Abt & B. F. Reiss (Eds.), Progress in clinical psychology. New York: Grune & Statton.

Wilkinson, R. T. (1969). Some factors influencing the effect of environmental stressors upon performance. Psychological Bulletin, 72, 260-272.

6. Is the task monotonous (the same response required to the same stimuli) or otherwise conducive to sleep?

Johnson, L. C. (1982). Sleep deprivation and performance. In W. B. Webb (Ed.) Biological Rhythms, Sleep, and Performance. New York: Wiley.

Mullaney, D. J., Fleck, P. A., Okudiaira, N., & Kripke, D. F. (1985). An automated system for administering continuous workload and for measuring sustained continuous performance. Behavior Research Methods, Instruments, & Computers, 17, 16-18.

Mullaney, D. J., Kripke, D. F., Fleck, P. A., & Johnson, L. C. (1983). Sleep loss and nap effects on sustained continuous performance. Psychophysiology, 20, 643-651.

Mullaney, D. J., Kripke, D. F., Fleck, P. A., & Okudaira, N. (1983). Effects of sustained continuous performance on subjects working alone and in pairs. Perceptual and Motor Skills, 57, 819-832.

Williams, H. L., Kearney, O. F., & Lubin, A. (1965). Signal uncertainty and sleep loss. Journal of Experimental Psychology, 69, 401-407.

7. **Is feedback or some other incentive used to motivate performers to try harder or persist longer at the task?**

Alluisi, E. A., & Childs, W. D. (1967). Sustained performance, work-rest scheduling, and diurnal rhythms in man. Acta Psychologica, 27, 436-442.

Haslam, D. R. (1983). The incentive effect and sleep deprivation. Sleep, 6, 362-368.

Johnson, L. C. (1982). Sleep deprivation and performance. In W. B. Webb (Ed.) Biological Rhythms, Sleep, and Performance. New York: Wiley.

Kjellberg, A. (1977). Sleep deprivation and some aspects of performance: Motivation, comment, and conclusions. Walking and Sleeping, 1, 133-135.

Naitoh, P., & Townsend, R. E. (1970). The role of sleep deprivation research in human factors. Human Factors, 12, 525-585.

Wilkinson, R. T. (1961). Interaction of lack of sleep with knowledge of results, repeated testing, and individual differences. Journal of Experimental Psychology, 62, 263-271.

Wilkinson, R. T. (1964). Effects of up to 60 hours of sleep deprivation on different types of work. Ergonomics, 17, 175-186.

Williams, H. L., Lubin, A., & Goodnow, J. J. (1959). Impaired performance with acute sleep loss. Psychological Monographs: General & Applied, 73, 1-26.

8. Is the task environment conducive to sleep?

Englund, C. E., Ryman, D. H., Naitoh, P., & Hodgdon, J. A. (1985). Cognitive performance during successive sustained physical work episodes. Behavior Research Methods, Instruments & Computers, 17, 75-85.

Johnson, L. C. (1982). Sleep deprivation and performance. In W. B. Webb (Ed.) Biological Rhythms, Sleep, and Performance. New York: Wiley.

Kleitman, N. (1923). On the effects of the loss of sleep. American Journal of Physiology, 67, 67-92.

Lee, M., & Kleitman, N. (1923). On the effects of the loss of sleep. American Journal of Physiology, 67, 141-152.

Wilkinson, R. T. (1963). Interaction of noise with knowledge of results, repeated testing, and individual differences. Journal of Experimental Psychology, 66, 332-337.

Part II

10. Relative to other tasks (critical or otherwise), how important is this task?

11. Relative to other tasks, how much does the performance of this task depend strictly on personal initiative?

12. Relative to other tasks, how much is the non-performance of this task likely to : (a) jeopardize human safety, (b) threaten mission outcome, or (c) cause costly equipment malfunction or delays?

Banderet, L. E., Stokes, J. W., Francesconi, R., Kowal, D. M., & Naitoh, P. (1981). Artillery teams in simulated sustained combat: Performance and other measures. In L. C. Johnson, D. I. Tepas, W. F. Colquhoun, & M. J. Colligan (Eds.), <u>The twenty-four hour workday: Proceedings of a symposium on variations in work-sleep schedules</u> (DHHS NIJOSH Pub. No. 81-127). Cincinnati, OH: U. S. Department of Health and Human Services, National Institute for Occupational Safety and Health.

Thorne, D., Genser, S., Sing, H., & Hegge, F. (1983, May 2-4). Plumbing human performance limits during 72 hours of high task load. In <u>Defense Group Proceedings: The human as a limiting element in military systems</u> (Vol.1, DS/A/DR[83]170). Toronto: NATO Defense Research Group.

Williams, H. L., & Lubin, A. (1967). Speeded addition and sleep loss. <u>Journal of Experimental Psychology</u>, <u>73</u>, 313-317.

Williams, H. L., Lubin, A., & Goodnow, J. J. (1959). Impaired performance with acute sleep loss. <u>Psychological Monographs: General & Applied</u>, <u>73</u>, 1-26.

APPENDIX E

SLEEP LOSS PERFORMANCE DATA

Experiment	Task	Task Rating	Hours w/o Sleep	% Baseline
1. Weiskotten & Ferguson (1930)	Ball tossing (Percent hits relative to controls)	215	15	100
			21	100
			27	100
			33	100
			39	100
			45	100
			51	100
			57	100
			63	100

Experiment	Task	Task Rating	Hours w/o Sleep	% Baseline
2. Weiskotten & Ferguson (1930)	Converting letters to telegraphic code (# of letters transposed in 5 minutes)	110	15	100
			21	97
			27	100
			33	98
			39	91
			45	71
			51	85
			57	93
			63	67

Experiment	Task	Task Rating	Hours w/o Sleep	% Baseline
3. Heslegrave & Angus (1985)	Simple iterative subtraction task (# correct responses/ minutes)	80	0 - 18	100
			23	96
			29	90
			35	90
			41	77
			47	52
			53	52

Experiment	Task	Task Rating	Hours w/o Sleep	% Baseline
4. Angus & Heslegrave	Message processing task (Message processing time in sec)	70	0 - 18	100
			20	87
			23	77
			26	77
			29	86
			32	77
			35	81
			41	76
			44	67
			47	52
			50	61
			53	61

Experiment	Task	Task Rating	Hours w/o Sleep	% Baseline
5. Williams, Lubin, & Goodnow	Choice reaction task (Reaction time in sec)	55	0	100
			30	89
			54	70
			69	53
			78	47
6. Angus & Heslegrave (1985)	Question processing task (Decode Questions) (Question processing time in sec)	50	0 - 18	100
			20	91
			23	70
			26	66
			29	66
			32	84
			35	68
			41	70
			44	45
			47	35
			50	54
			53	43

Experiment	Task	Task Rating	Hours w/o Sleep	% Baseline
7. Williams, Lubin, & Goodnow (1959)	Memory span task (# items recalled)	55	0	100
			27	78
			51	48
			75	13
8. Angus & Heslegrave (1985)	Encoding/decoding task (# of responses per min)	50	0 - 18	100
			22	67
			28	75
			34	73
			40	73
			46	42
			52	47

Experiment	Task	Task Rating	Hours w/o Sleep	% Baseline
9. Angus & Heslegrave (1985)	Vigilance task (% correct)	45	0 - 18	100
			19	88
			25	71
			31	69
			37	72
			43	59
			49	39
			55	65
10. Angus & Heslegrave (1985)	Serial reaction task (# of responses/min)	40	0 - 18	100
			22	76
			28	75
			34	71
			40	83
			46	40
			52	48

Experiment	Task	Task Rating	Hours w/o Sleep	% Baseline
11. Williams, Lubin, & Goodnow (1959)	Vigilance (visual) (# errors of commission)	35	0	100
			28	50
			52	25
			76	20
12. Williams, Lubin, & Goodnow (1959)	Vigilance (auditory) (# errors of commission)	35	0	100
			28	67
			52	25
			76	20
13. Angus & Heslegrave (1985)	Logical reasoning task (# of correct responses/min)	30	0 - 18	100
			22	59
			28	61
			34	62
			40	58
			46	28
			52	42

Experiment	Task	Task Rating	Hours w/o Sleep	% Baseline
14. Thorne, et al., (1983)	2-Letter search (% correct/mean time)	50	0 - 8	100
			24	96
			30	90
			36	94
			42	80
			48	62
			54	70
			60	65
			66	50
			72	45

Experiment	Task	Task Rating	Hours w/o Sleep	% Baseline
15. Thorne, *et al.*, (1983)	6-Letter search (% correct/mean time)	40	0 - 8	100
			24	94
			30	80
			36	88
			42	82
			48	50
			54	70
			60	60
			66	55
			72	65

Experiment	Task	Task Rating	Hours w/o Sleep	% Baseline
16. Thorne, _et al_., (1983)	Two-column addition (% correct/ mean time)	50	0 - 8	100
			24	88
			30	88
			36	88
			42	75
			48	45
			54	80
			60	60
			66	55
			72	45

Experiment	Task	Task Rating	Hours w/o Sleep	% Baseline
17. Thorne, *et al.*, (1983)	Logical reasoning (% correct/ mean time)	40	0 - 8	100
			24	88
			30	80
			36	90
			42	70
			48	58
			54	60
			60	35
			66	40
			72	30

Experiment	Task	Task Rating	Hours w/o Sleep	% Baseline
18. Thorne, *et al.*, (1983)	Digit recall (% correct/ mean time)	40	0 - 8	100
			24	90
			30	90
			36	85
			42	82
			48	55
			54	57
			60	55
			66	35
			72	35

Experiment	Task	Task Rating	Hours w/o Sleep	% Baseline
19. Thorne, *et al.*, (1983)	Serial add/subtract (% correct/mean time)	40	0 - 8	100
			24	80
			30	82
			36	80
			42	55
			48	45
			54	35
			60	35
			66	30
			72	25

Experiment	Task	Task Rating	Hours w/o Sleep	% Baseline
20. Thorne, *et al*., (1983)	Pattern Recognition I (% correct/mean time)	50	0 - 8	100
			24	88
			30	80
			36	120
			42	75
			48	45
			54	50
			60	45
			66	35
			72	48

Experiment	Task	Task Rating	Hours w/o Sleep	% Baseline
21. Thorne, *et al.*, (1983)	Pattern Recognition II (% correct/mean time)	40	0 - 8	100
			24	115
			30	70
			36	94
			42	75
			48	42
			54	60
			60	45
			66	30
			72	22

APPENDIX F

Multiple Regression Analysis

The multiple regression model used to analyze the data is as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$$

where:

Y = % Baseline (Dependent Variable)

β_0 = Intercept

β_1 = Hours without Sleep (Independent Variable)

β_2 = Task Rating (Independent Variable)

e = Error

and

β_0 = 96.7163

β_1 = -0.9579

β_2 = 0.2057

The statistical package used to generate the analysis is the Statview™ statistical package for the Macintosh computer.

Multiple - Y : Baseline Two X variables

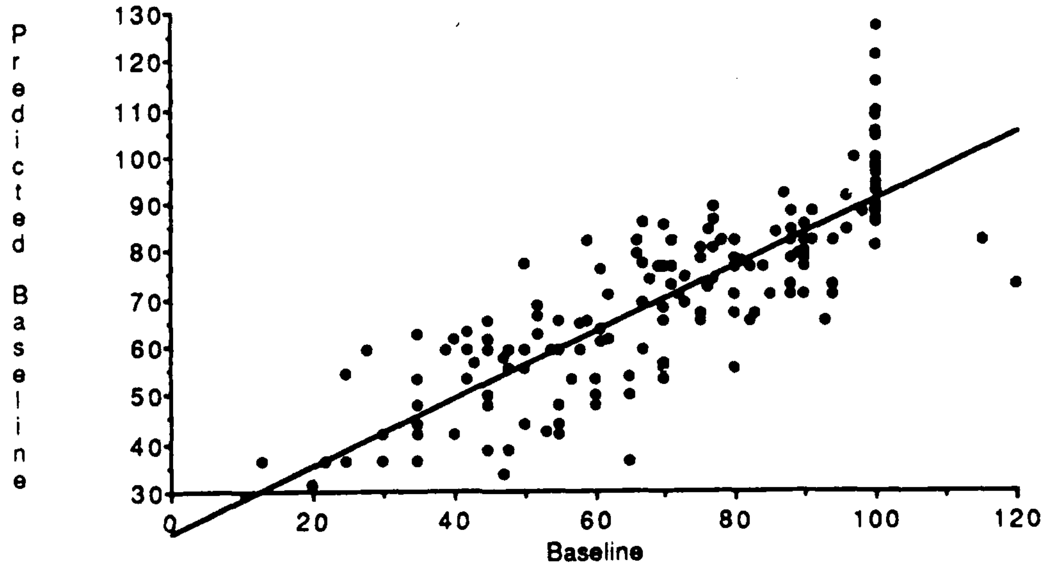
DF:	R-squared:	Std. Err.:	Coef. Var.:
174	.69529862	13.07893361	18.51491167

Analysis of Variance Table

Source	DF:	Sum Squares:	Mean Square:	F-test:
REGRESSION	2	6.71382573E4	3.35691286E4	196.24355288
RESIDUAL	172	2.94220627E4	171.05850429	p ≤ .0001
TOTAL	174	96560.32		

Beta Coefficient Table

Parameter:	Value:	Std. Err.:	T-Value:	Partial F:
INTERCEPT	96.7163473	2.83710548	34.0897961	
NS	-.95793196	.0542922	-17.64400778	311.31101069
Rating	.20570088	.02457125	8.3716082	70.08382393



PREDICTED VALUES

OBS NO.	BASELINE	NS	RATING	PREDICTED VALUES
1	100	15	215	126.573
2	100	21	215	120.825
3	100	27	215	115.078
4	100	33	215	109.33
5	100	39	215	103.583
6	100	45	215	97.835
7	100	51	215	92.088
8	100	57	215	86.34
9	100	63	215	80.592
10	100	15	110	104.974
11	97	21	110	99.227
12	100	27	110	93.479
13	98	33	110	87.732
14	91	39	110	81.984
15	71	45	110	76.237
16	85	51	110	70.489
17	93	57	110	64.741
18	67	63	110	58.994
19	100	18	80	95.93
20	96	23	80	91.14
21	90	29	80	85.392
22	90	35	80	79.645
23	77	41	80	73.897
24	52	47	80	68.15
25	52	53	80	62.402
26	100	18	70	93.873
27	87	20	70	91.957
28	77	23	70	89.083
29	77	26	70	86.209
30	86	29	70	83.335
31	77	32	70	80.462
32	81	35	70	77.588
33	76	41	70	71.84
34	67	44	70	68.966
35	52	47	70	66.093
36	61	50	70	63.219
37	61	53	70	60.345
38	100	0	55	108.03
39	89	30	55	79.292
40	70	54	55	56.302
41	53	69	55	41.933
42	47	78	55	33.311
43	100	18	50	89.759
44	91	20	50	87.843

PREDICTED VALUES

OBS NO.	BASELINE	NS	RATING	PREDICTED VALUES
45	70	23	50	84.969
46	66	26	50	82.095
47	66	29	50	79.221
48	84	32	50	76.348
49	68	35	50	73.474
50	70	41	50	67.726
51	45	44	50	64.852
52	35	47	50	61.979
53	54	50	50	59.105
54	43	53	50	56.231
55	100	0	55	108.03
56	78	27	55	82.166
57	48	51	55	59.175
58	13	75	55	36.185
59	100	18	50	89.759
60	67	22	50	85.927
61	75	28	50	80.179
62	73	34	50	74.432
63	73	40	50	68.684
64	42	46	50	62.937
65	47	52	50	57.189
66	100	18	45	88.73
67	88	19	45	87.772
68	71	25	45	82.025
69	69	31	45	76.277
70	72	37	45	70.529
71	59	43	45	64.782
72	39	49	45	59.034
73	65	55	45	53.287
74	100	18	40	87.702
75	76	22	40	83.87
76	75	28	40	78.122
77	71	34	40	72.375
78	83	40	40	66.627
79	40	46	40	60.88
80	48	52	40	55.132
81	100	0	35	103.916
82	50	28	35	77.094
83	25	52	35	54.103
84	20	76	35	31.113
85	100	0	35	103.916
86	67	28	35	77.094
87	25	52	35	54.103
88	20	76	35	31.113

PREDICTED VALUES

OBS NO.	BASELINE	NS	RATING	PREDICTED VALUES
89	100	18	30	85.645
90	59	22	30	81.813
91	61	28	30	76.065
92	62	34	30	70.318
93	58	40	30	64.57
94	28	46	30	58.823
95	42	52	30	53.075
96	100	8	50	99.338
97	96	24	50	84.011
98	90	30	50	78.263
99	94	36	50	72.516
100	80	42	50	66.768
101	62	48	50	61.021
102	70	54	50	55.273
103	65	60	50	49.525
104	50	66	50	43.778
105	45	72	50	38.03
106	100	8	40	97.281
107	94	24	40	81.954
108	80	30	40	76.206
109	88	36	40	70.459
110	82	42	40	64.711
111	50	48	40	58.964
112	70	54	40	53.216
113	60	60	40	47.468
114	55	66	40	41.721
115	65	72	40	35.973
116	100	8	50	99.338
117	88	24	50	84.011
118	88	30	50	78.263
119	88	36	50	72.516
120	75	42	50	66.768
121	45	48	50	61.021
122	80	54	50	55.273
123	60	60	50	49.525
124	55	66	50	43.778
125	45	72	50	38.03
126	100	8	40	97.281
127	88	24	40	81.954
128	80	30	40	76.206
129	90	36	40	70.459
130	70	42	40	64.711
131	58	48	40	58.964
132	60	54	40	53.216

PREDICTED VALUES

OBS NO.	BASELINE	NS	RATING	PREDICTED VALUES
133	35	60	40	47.468
134	40	66	40	41.721
135	30	72	40	35.973
136	100	8	40	97.281
137	90	24	40	81.954
138	90	30	40	76.206
139	85	36	40	70.459
140	82	42	40	64.711
141	55	48	40	58.964
142	57	54	40	53.216
143	55	60	40	47.468
144	35	66	40	41.721
145	35	72	40	35.973
146	100	8	40	97.281
147	80	24	40	81.954
148	82	30	40	76.206
149	80	36	40	70.459
150	55	42	40	64.711
151	45	48	40	58.964
152	35	54	40	53.216
153	35	60	40	47.468
154	30	66	40	41.721
155	25	72	40	35.973
156	100	8	50	99.338
157	88	24	50	84.011
158	80	30	50	78.263
159	120	36	50	72.516
160	75	42	50	66.768
161	45	48	50	61.021
162	50	54	50	55.273
163	45	60	50	49.525
164	35	66	50	43.778
165	48	72	50	38.03
166	100	8	40	97.281
167	115	24	40	81.954
168	70	30	40	76.206
169	94	36	40	70.459
170	75	42	40	64.711
171	42	48	40	58.964
172	60	54	40	53.216
173	45	60	40	47.468
174	30	66	40	41.721
175	22	72	40	35.973

RESPONSES TO STAFFING CONCERNS

Question a: Provide the multiple regression equation (b weights and constant) which were used to develop Table 3 (Sleep Loss Effects Prediction Matrix). This information is needed so that we can use the equation to predict soldier performance with maximum precision based on instrument ratings and hours without sleep.

Answer: The Beta Coefficient Table for the multiple regression is reproduced below; it has been included in Appendix F of the final report.

Beta Coefficient Table

Parameter:	Value:	Std. Err.:	T-Value:	Partial F:
INTERCEPT	96.7163473	2.83710548	34.0897961	
NS	-.95793196	.0542922	-17.64400778	311.31101069
Rating	.20570088	.02457125	8.3716082	70.08382393

The Statview™ statistical package for the Macintosh computer was used for the regression.

Additional study resulted in the use of two independent variables, (Hours Without Sleep and Task Rating). The product of Hours Without Sleep and Task Rating depicting interactive effects between the independent variables was deleted from the final report after further study.

Question b: Table 3 presents only per cent of baseline performance in view of survey ratings and hours without sleep. Could the methodology be altered to show both time to complete task and precision with which task is completed? For example, sleep loss might result in some tasks being performed much slower, although not necessarily less competently than baseline. Is there a way to break these two performance dimensions out separately?

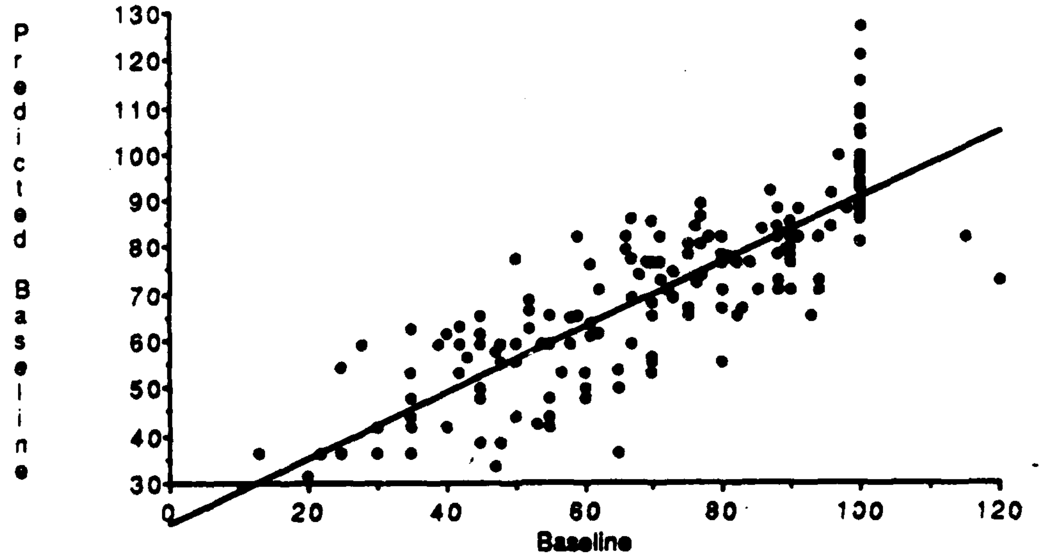
> **Answer**: Insofar as data can be found, the methodology could be used to estimate the effects of human factors on the accuracy of tasks according to their task ratings. A more difficult problem arises in applying degraded precision to the combat models. For some model parameters, simple error rates, or probability of failing to perform the task "satisfactorily", might be useful. If a task were not performed at all, or if it were performed so poorly that the system completely failed to perform its function, then the outcome with respect to model activities could be determined. For example, a target would either be hit or missed. It is less clear how to represent degrees of accuracy and how they affect model inputs and outcomes. For example, what is the effect of a somewhat less than perfect sight picture on the ability to hit and kill a target? Data to support the determination of effects on precision were not found during this study. Inputs could, of course, be generated using subject matter expertise; this would be more speculative than data derived from experimental research. Research into the relationship between individual task performance levels and overall system performance is known to be taking place, and more will undoubtedly be undertaken. When results from these efforts become available, they may be helpful in establishing empirical relationships between individual task performance precision and combat model inputs.

Question c: How much unique variance was explained by each individual predictor in the multiple regression equation (i.e., how much was accounted for by each variable after all the others are already in the predictive equation)?

> **Answer**: See the Beta Coefficient Table in the answer to Question a above.

Question d: When talking about tentative validity, it might be helpful to plot actual versus predicted curves of performance degradation (per cent of baseline).

Answer: A plot of actual versus predicted curves of performance degradation (percent of baseline) is depicted below and also found on page 75, Appendix F.



Question e: A total of 35 unit/systems variables were found to be possibly influenced by soldier performance. What are these variables?

Answer: Through further study and research, we reduced this number to 16 variables. They are listed in Appendix A of the final report.

Question f: Question 5 on the survey instrument concerns time -- and yet the definition of "mental effort load" (page 11, which item 5 is measuring along with items 1 and 2) says it is free of time considerations.

> **Answer**: We deleted the words "without regard to time" and changed the sentence to read: "Mental effort load depends on the absolute amount of attentional capacity or effort required by the task and the duration of the task." Question 5 does not relate to time in the sense of "time pressure" which imposes its own unique demands (Time Load). Question 5 relates to time in the sense of task duration.

Question g: The example on page 16 provides an incorrect prediction based on the explanation of Table 3 on the bottom of page 15. Are numbers the percent who can or cannot perform?

> **Answer**: We corrected the wording in the text. The numbers represent the percent who are expected to be able to perform a task correctly.

Question h: I assume the use of the product of task ratings and hours without sleep was a move in the direction of assessing the interactive effects of these variables on per cent of baseline performance. If so, more explanation should be given to these results -- and/or any other efforts which were directed toward assessment of interaction effects.

> **Answer**: The use of the product of task ratings and hours without sleep as a third independent variable was intended to explore the interactions between the two. Further examination of the regression has resulted in the elimination of the product as an independent variable (See answers to Questions a and c above).

Question i: The rating scale method explanation is too vague. It is not explained in sufficient detail that a government employee could do it. Revision should include an outline in step-by-step fashion which very clearly illustrates this procedure from beginning to end. An example would also prove very useful here.

> **Answer**: As suggested by Question t, we have moved the section on the Rating Scale Method to an appendix.. This method offers one potential solution to problems associated with estimating the joint effects of multiple, interacting variables. For this reason, we wanted readers to have some familiarity with the method.

Question j: A section has been omitted just prior to the top of page 25.

> **Answer**: This problem has been corrected in the final report.

Question k: Item 2 on the rating scale rates soldier tasks from lots of uncertainty and logical reasoning to great certainty and physical strength. Can a task which requires mostly strength ever have an outcome with lots of uncertainty? If so, this item would not permit this combination of rating.

> **Answer**: The Task Rating Sheet has been revised to reflect this rationale in the final report. Our intent was to suggest a task involving essentially no mental demands. A simple strength task like squeezing a hand dynamometer appeared about as far down this end of the continuum as we could get.

Question l: I was disappointed that this methodology is limited to consideration of only a few human factors variables, such as sleep loss and task requirements (physical vs. mental, etc.). What is needed is a comprehensive study which investigates the impact that numerous human factors have on soldier performance when all are impinging simultaneously on the unit. (Lack of available, properly-collected data is a major problem here.) Nevertheless, this study provides a major step forward in the provision of a methodology that may someday be expanded to include additional human factors.

Answer: We understand your disappointment, but we also hope you can appreciate the difficulties associated with dealing with human performance variables. As noted earlier, the important limitations are not in the methodologies that we proposed. These methodologies are flexible enough to account for the effects of variables other than sleep loss. They also appear flexible enough to account for the effects of several variables in combination with one another. The primary limitation is in what we know about the effects of different variables on human performance. The more of this knowledge that we can get, the better our proposed methodologies should work. Of course, there is nothing to prevent us from moving beyond the data. But, in moving beyond the data, or speculating, about the effects of particular variables on human performance, every effort must be made to develop clear, testable rationales for our decisions and to document these decisions appropriately.

Question m. CASTFOREM should have been used (instead of VIC) because it comes closer to portraying what soldiers really do in combat.

Answer. VIC appeared a reasonable point of departure for this effort, but there is no reason not to use CASTFOREM in subsequent investigations and expansion on the process.

Question n. This methodology will not work with VIC because an aggregated algorithm that follows adjusting the inputs (the Lanchester equations) smothers the effects that human factors have on performance.

Answer. It is understood that VIC aggregates detailed inputs for the application of the Lanchester equations. However, since the methodology would adjust existing VIC inputs to account for human factors variables, then the impacts of human factors should be visible in the results. In other words, the model would be as responsive to changes in performance inputs derived from human factors considerations as it would be to changes in performance derived from engineering changes or other materiel-related considerations.

Question o. Additional human factors need to be added to the inventory.

Answer. We agree, and we believe we now have a methodology that will allow us to proceed in this direction. We just wanted to be sure we could walk before we tried to run.

Question p. Interactive effects and inter-rater reliability need to be more thoroughly explored.

Answer. Again we agree.

Question q. Other models (JANUS or CASTFOREM) would provide a clearer, unaggregated picture for soldier performance modeling than VIC. Because VIC is a division/corps model, it will not realistically portray effects of soldier performance degradation.

Answer. See the answers to Questions m and n. It is agreed that other models could provide for more direct means of examining human factors effects. However, changes to VIC inputs attributable to human factors would be as visible as changes attributable to other considerations, such as alternative system designs. Once the input changes are made, their sources are invisible to the model algorithm and outputs.

Question r. The report does not provide validation results which are consistent with the Army definition as this concept pertains to models: "A process of determining that (a model) is an accurate representation of the intended real-world entity from the perspective of its intended use.

Answer. We agree. We have defined an approach which appears to have some potential for improving the fidelity of our combat models, and we believe that it deserves some further consideration and testing. However, at this point, none of us can assure that the approach will yield valid or reliable results.

Question s. The Performance Prediction Table introduces considerable confusion into the methodology. Several reviewers thought it was a look-up table to be used in the final methodology. If it is essential in terms of explaining the derivation of the methodology, recommend this table be relegated to an appendix.

Answer. This section has been rewritten.

Question t. The Rating Scale Method was "tacked" onto the report as an alternative to be used. As such, it detracted from the flow and logic of the main methodology. Recommend it be relegated to an appendix.

Answer. This is a good recommendation, and we have followed it.

Question u. The rating scale method was not described in sufficient detail that it could be used. The pros and cons of each is needed to help individuals determine which to use in different situations. Also, if two different methodologies are presented, they should be tracked together throughout all remaining portions of the report.

Answer. We have provided additional detail on the Rating Scale Method, but we decided to move information on this method to an appendix. This method holds potential for application in the area, but it is not our method of first choice.

Question v. The survey ignores a very important effect of sleep loss -- nonperformance. This scale should be coordinated with SMEs such as COL Greg Belenky (on TAG), COL Krueger (on TAG), or MAJ Lew of WRAIR to ensure that is measures the most important effects of sleep loss.

Answer. Information on Part II of the survey, which is designed to deal with the issue of nonperformance,was inadvertently omitted from the draft report. This information has been included in the final report. Also, the draft report has been coordinated with COL Belenky and COL Krueger.

Question w. A more generic rating scale is needed -- perhaps called a Combat Stress Effects Scale -- so that the effects of sleep loss could be compared with other types of stressors, such as fatigue, noise, cold, heat, vibrations, etc. A "sleep loss" scale is too specific to be widely used by the military community.

> **Answer.** We agree. We even like the name "Combat Stress Effects Scale"! We did not mean to suggest that a sleep loss scale alone would accommodate the needs of the combat modeling community. We know more is needed to model human performance -- much more. Our focus on the sleep loss research was for demonstration purposes only. Other areas could have been considered as well, depending primarily on the availability of the data. Thus, at this point, we recognize the limited scope of the demonstration, but still believe that we have identified a method which holds real potential for expansion and use in the modeling arena.

Question x. A significant limitation of any methodology which attempts to adjust soldier performance in view of human factor variables is that soldier effects are confounded by attempts by military organizations to limit the adverse impact of said variables on performance.

> **Answer.** When the military attempts to adjust soldier performance, it uses whatever means it has available -- training, leadership, cohesion, incentives, work-rest schedules, etc. -- to do so. We can employ the proposed methodology to predict the effects of these variables in the same way that we used it to predict the effects of sleep loss. This is not a limitation of the methodology. This is another problem which stems from our relative lack of understanding of the effects of human performance variables, either in isolation or in combination with one another.

Question y. This method attempts to account for the effect of human factors on individuals. However, the methodology does not demonstrate how to roll this effect up to account for performance changes for crews, units, or large forces. Here you should bear in mind that a single crew member might be very tired without degrading the combat performance of a tank.

Answer. The methodology can easily be adapted to account for the performance of individuals in crews, units, and large forces. It is only a matter of modifying the questions and weighting the response alternatives appropriately. Again, the problem does not lie with the methodology. The problem lies in the fact that we lack much needed information about the behavior of soldiers in crews, units, or large forces. We do not know how to modify the questions and weight the response alternatives appropriately. And, we probably never will know everything that we need to know. At some point, we will have to be content to make some educated guesses.

Question z. The specific application to VIC target acquisition (page 7) might not translate well to other models, such as CASTFOREM.

Answer. We believe you are correct. Some consideration would have to be given to models of interest on a case-by-case basis.

Question aa. The example (page 16) which refers to a task to "identify and employ hand grenades" has little relevance in the context of VIC. In the interests of credibility, another example should be used.

Answer. This a good point. The text has been altered accordingly.

Question bb. The method of conjoint analysis is limited by the degree of agreement about the relative levels of the three tasks mentioned. Lack of consensus renders this method useless.

Answer. This question appears to concern inter-rater reliability, which has not been a problem, at least when tests have been conducted using the Subjective Workload Assessment Technique (SWAT). Correlations reflecting inter-rater reliability have consistently been high and positive (e.g., 0.70 - 0.90) (e.g., Reid, Shingledecker, & Eggemeier, 1981).

Question cc. The multiple regression methodology needs to be described in greater detail. What practical (as opposed to statistically) significance is evidenced by results?

Answer. See Question c.

Question dd. Army Research Institute published a report which considered the sensitivity of VIC to human factor variables.

Answer. The ARI report has much to offer but because of time and resource constraints, we could not incorporate its results into this report.