Technical Report 599

# A Model of Mission Accomplishment in Simulated Battle

Gary S. Thomas and Thomas G. Cocklin

ARI Field Unit at Fort Leavenworth, Kansas
**Systems Research Laboratory**



U. S. Army

Research Institute for the Behavioral and Social Sciences

November 1983

# U. S. ARMY RESEARCH INSTITUTE
# FOR THE BEHAVIORAL AND SOCIAL SCIENCES

## A Field Operating Agency under the Jurisdiction of the
## Deputy Chief of Staff for Personnel


EDGAR M. JOHNSON
Technical Director

L. NEALE COSBY
Colonel, IN
Commander


Technical review by

Jay S. Coke
Ira T. Kaplan

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|
| 1. REPORT NUMBER: Technical Report 599 | 2. GOVT ACCESSION NO.: AD-A148 234 |
| | 3. RECIPIENT'S CATALOG NUMBER: |
| 4. TITLE (and Subtitle): A MODEL OF MISSION ACCOMPLISHMENT IN SIMULATED BATTLE | 5. TYPE OF REPORT & PERIOD COVERED: Final |
| | 6. PERFORMING ORG. REPORT NUMBER: -- |
| 7. AUTHOR(s): Gary S. Thomas, Thomas G. Cocklin | 8. CONTRACT OR GRANT NUMBER(s): -- |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS: U.S. Army Research Institute for the Behavioral and Social Sciences, 5001 Eisenhower Avenue, Alexandria, VA 22333-5600 | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 2Q263744A795 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS: -- | 12. REPORT DATE: November 1983 |
| | 13. NUMBER OF PAGES: 28 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office): -- | 15. SECURITY CLASS. (of this report): UNCLASSIFIED |
| | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE: -- |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

--

18. SUPPLEMENTARY NOTES

A substantial portion of this report appears in the Proceedings of the Human Factors Society, 27th Annual Meeting, Norfolk, VA, 1983.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

| | |
|---|---|
| Battlefield performance | Performance evaluation |
| Battle simulation | Policy capturing |
| Command and control training | Decision making |
| Performance prediction | Military expert judgments |
| Mission accomplishment | |

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

The purpose of this research was to formulate a composite measure of mission accomplishment that could be used to assess simulated combat performance of battalion command groups, training command and control procedures, in a computer-driven battle simulation. Mission accomplishment was defined in terms of six measurable components or objectives of a covering force/delay mission. Four retired Army officers served as judges for all experiments. Experiment I demonstrated that two of the objectives could be -- (Continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Item 20 (Continued)

meaningfully combined into a single measure that was used in previous research as an index of relative losses by opposing forces.

Multiple regression was used in Experiment II to describe how the judges assigned mission accomplishment scores to 216 hypothetical battle outcomes including measures of each mission objective. Experiment III compared judges' assessments of 10 actual battle outcomes to the mission accomplishment scores predicted by the regression models derived in Experiment II. Inter-rater agreement was high in the assessments of actual battle outcomes, and individual predictor models accounted for more than 94% of the variance in ratings of these data for three of four judges. A composite model of mission accomplishment accounted for 93% of the variance in the average assessments of all four judges. Results indicate that it may be feasible to replace the expert judges with the composite model of mission accomplishment.

Technical Report 599

# A Model of Mission Accomplishment in Simulated Battle

Gary S. Thomas and Thomas G. Cocklin

**Submitted by**
Robert S. Andrews, Chief
**ARI Field Unit at Fort Leavenworth, Kansas**

**Approved as technically adequate and submitted for publication by**
Jerrold M. Levine, Director
**Systems Research Laboratory**

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

November 1983

**Army Project Number**
**2Q263744A795**

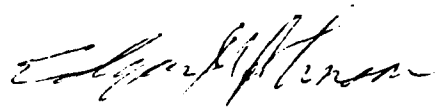
**Training and Simulation**

ARI Research Reports and Technical Reports are intended for sponsors of R&D tasks and for other research and military agencies. Any findings ready for implementation at the time of publication are presented in the last part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

---

## FORWORD

The U.S. Army Research Institute, Fort Leavenworth Field Unit conducts a systems and training research program in support of the Combined Arms Center (CAC). The current research supports an on-going effort to develop measures of simulated battlefield performance of battalions, training command and control procedures, in the Combined Arms Tactical Training Simulator, the prototype for the Army Training Battle Simulation System (ARTBASS). This research describes a procedure for characterizing performance on the simulated battlefield, to which measures of the command and control process can be compared.



EDGAR M. JOHNSON
Technical Director

ACKNOWLEDGMENTS

The authors would like to thank Dr. David Thissen and Ms. Janet Marquis from the University of Kansas for their statistical consultation.

A MODEL OF MISSION ACCOMPLISHMENT IN SIMULATED BATTLE

## EXECUTIVE SUMMARY

### Objective:

The purpose of this research was to determine the feasibility of formulating a composite measure of mission accomplishment that included measures of important mission objectives. Such a composite measure could be used in place of expert judges to assess battlefield performance in computer-driven battle simulations, such as CATTS/ARTBASS. The composite measure of battlefield performance could be used in subsequent research in comparison to measures of command and control processes.

### Procedures:

Four retired Army officers served as military expert judges due to their extensive experience in combat and/or combat modeling. Their ranks ranged from lieutenant colonel to brigadier general. Judges rated the degree of mission accomplishment of 216 hypothetical battle outcomes which included measures of covering force mission objectives: (1) attrit the enemy, (2) minimize friendly losses, (3) remain combat effective, (4) delay the OPFOR as far forward of the MBA as possible, (5) gather intelligence as to the enemy's strength and likely courses of action, and (6) be prepared to conduct passage of lines. Based on the findings in Experiment I, measures of objectives 1 and 2 were combined into a single measure of relative losses. Ratings were analyzed by multiple regression to mathematically describe judges decision rules in assigning mission accomplishment scores to the hypothetical battle outcomes. Judges later rated 10 battle outcomes from actual CATTS exercises in terms of mission accomplishment, and these scores were compared to those predicted by the regression models.

### Findings:

1. Regression models representing judges' decision rules, accounted for more than 96% of the variance in the mission accomplishment scores for hypothetical data. This indicated that regression models accurately described judges behavior.

2. Three of four judges were able to reapply their decision rules to actual CATTS data as indicated by $R^2$ in excess of .94 between mission accomplishment scores predicted by models and scores provided by judges.

3. There was high inter-rater agreement on mission accomplishment scores for actual battle outcomes, eventhough there were differences in judges' decision rules.

4. A composite model of mission accomplishment was formulated which accurately predicted average mission accomplishment scores provided by judges.

Utilization of Findings:

It appears possible to capture decision policies of military experts in the form of a multiple regression model. The model could be used in place of judges to assess mission accomplishment in future command and control training research with systems such as CATTS. The measure of mission accomplishment formulated in this research appears to be a more comprehensive measure than simple loss exchange ratios.

A MODEL OF MISSION ACCOMPLISHMENT IN SIMULATED BATTLE

CONTENTS

LIST OF TABLES

# A Model of Mission Accomplishment in Simulated Battle

Gary S. Thomas
and
Thomas G. Cocklin

## INTRODUCTION

Current Army doctrine relies heavily on command and control ($C^2$) to insure success on the modern battlefield. Command staffs will be required to exercise exceptional $C^2$ to succeed on the highly volatile, complex, and lethal battlefield, especially when fighting outnumbered. It is, therefore, necessary for the Army to train command groups in the performance of $C^2$ processes and behaviors to increase the likelihood of their survival. In recent years, battle simulations have gained increased credibility as systems for training $C^2$ processes at tactical echelons from battalion through corps. Of the battle simulations developed to support such training, the most sophisticated in terms of the extent of automation and, therefore, the ability to represent battlefield events in real time, is the Combined Arms Tactical Training Simulator (CATTS).

CATTS is used to train battalion command groups (BCG) and serves as a test bed to improve training procedures and to specify requirements for future simulations. The simulation provides a computer-driven, real-time, free-play exercise to train maneuver BCGs in the control and coordination of combined arms operations. CATTS simulates the actions of units in combat, moves elements on and about the battlefield, calculates intervisibility and detection between forces, calculates weapon-to-target ranges, and determines losses inflicted by individual weapon systems. The computer takes as input $C^2$ decisions made by the BCG and determines the consequences of those decisions in the form of movements and losses for each platoon modeled in the system.

Efforts have been directed at identifying and measuring components of the $C^2$ process that are trained in CATTS exercises (Barber and Kaplan, 1979; and Kaplan and Barber, 1979). These efforts resulted in a questionnaire based on the Army Training and Evaluation Program (ARTEP) that was intended to assess the ability of BCGs to perform certain critical subtasks inherent in the $C^2$ process. Subsequent research using this instrument indicated that ratings of ARTEP subtasks were subject to many of the problems associated with subjective ratings, such as limited inter-rater agreement, lack of item discriminability, rater bias, etc. (Thomas, Barber, and Kaplan, 1983).

Research has been conducted to develop more objective measures of BCG performance on the simulated battlefield modeled in CATTS. Thomas (1983) investigated battle simulation outcomes (e.g., mathematical relationships of friendly and enemy weapon losses) as potential measures of battle performance. These measures correlated very highly with overall ratings of ARTEP

performance when type of mission (attack vs. covering force) was controlled. These simulation outcomes were also responsive to manipulations in CATTS system characteristics. For example, Thomas et al. (1983) reported that combat ratio, weather, mission type, and reduced jamming of communication nets resulted in significant differences in simulation outcome scores. But, simulation outcomes did not correlate with ratings of performance.

Although the results are encouraging in terms of using simulation outcome scores as measures of BCG performance in simulated battle, these scores reflect only part of the battlefield mission objectives stated in CATTS exercises. Therefore, the current research attempted to develop a procedure for measuring relevant dimensions of battlefield performance (mission objectives) and to determine how these dimensions could be combined into a composite measure of overall mission accomplishment. Such a composite measure of performance should be more comprehensive than simple measures of relative losses; and therefore, more appropriate as an evaluation metric.

Procedures for determining how such multiple sources of information are combined into overall assessments by judges can be found in the decision making literature (see Slovic and Lichtenstein, 1971, for a review). A variety of studies have proposed the use of linear regression models to represent clinical judgment (Dudycha and Naylor, 1966; Goldberg, 1968; Hammond, Hirsch, and Todd, 1964; Hirsch, Hammond, and Hirsch, 1964; Hoffman, Slovic, and Rorer, 1968; Naylor and Wherry, 1965; Schenck and Naylor, 1968; Wherry and Naylor, 1966; Wiggins and Hoffman, 1968). For example, (Slovic, 1969) in a study of stockbrokers' judgments on corporate factors that predict fluctuations in stock index, concluded that these expert decisions were only linear and additive. These linear representations have been shown to realistically represent the decision rule of judges, and in fact, it has been demonstrated that the regression model has better predictive quality than the judges themselves (e.g., Meehl, 1954, 1965). A "bootstrapping" technique (as reviewed by Dawes and Corrigan, 1974) has been used to construct modeled representations of judges' decision rules. The validity of the model can then be tested against a new set of decision choices made by the same judge. Alternatively, a linear model can be constructed to represent all judges in general. A high level of prediction is typically demonstrated, and the composite model tends to be a better predictor of judges ratings than is any single model obtained from individual judges.

The current research used the least squares regression model to describe how military judges combined measures of several components of covering force mission performance into a single measure of overall mission accomplishment. Validity of the predictive models was assessed by comparing judges' ratings of actual CATTS data to ratings of mission accomplishment predicted by the models. In addition, a composite model representing all judges was formulated. Such a composite model has the potential of replacing judges in subsequent mission accomplishment evaluations.

## EXPERIMENT I

A preliminary experiment was conducted to determine if the simulation outcome measures used in previous research (Thomas, 1983; Thomas, et al., 1983) were meaningful mathematical combinations of friendly and enemy losses. If the simulation outcomes represent meaningful ways of combining the losses sustained by opposing forces, a high degree of relationship between military judges' ratings of the loss measures and the magnitude of the simulation outcomes, which are based on these loss data, should be expected.

## METHOD

### Subjects

Subjects were four retired military officers whose rank ranged from lieutenant colonel to brigadier general. These judges were selected because of their extensive experience in combat or in combat modeling. Judges were paid for participation.

### Stimulus Materials

Stimulus materials, which appear in Appendix A, were the percent of friendly forces surviving and the percent of opposing force (OPFOR) attrited in the simulated battle of CATTS exercises as reported in Thomas, et al., (1983). These measures were collected in eight attack and eight covering force missions, with two each performed by four BCGs. These stimulus materials are actually components of the simulation outcomes reported in the above research.

### Procedures

Judges were presented with all possible combinations (120) of the friendly surviving and OPFOR attrited data from the 16 exercises. Judges indicated which of the two in each pair reflected the most favorable battle outcomes from the perspective of friendly forces.

Simulation outcomes, which appear in Table 1, were calculated as follows: Relative exchange ratio (RER) equals the percentage of OPFOR weapon systems lost divided by the percentage of friendly weapons lost in battle. Surviving maneuver force ratio differential (SMFRD) equals the percent of friendly forces surviving minus the percent of OPFOR surviving. The change in combat ratio ( $\Delta$CR) equals beginning combat ratio minus end of battle combat ratio, and that divided by beginning combat ratio. Combat ratio equals total OPFOR divided by total friendly forces available for battle. The command and control index of lethality levels ($C^2$ILL) equals one-half the percent of friendly forces surviving plus the percent of OPFOR attrited.

Table 1

Calculations of Simulation Outcomes*

| | | |
|---|---|---|
| Relative Exchange Ratio | = | $\frac{\text{Percentage of OPFOR Lost}}{\text{Percentage of Friendly Forces Lost}}$ |
| Surviving Maneuver Force Ratio Differential | = | Percentage of Friendly Forces Surviving minus the Percentage of OPFOR Surviving |
| $C^2$ILL Ratio | = | 1/2 (Percentage of Friendly Forces Surviving) plus the Percentage of OPFOR Lost |
| ΔCR Combat Ratio | = | $\frac{\text{Initial Combat Ratio minus Ending Combat Ratio}}{\text{Initial Combat Ratio}}$ |

*All losses are based on ΣEW X ET per exercise where EW = equipment weighting factor and ET = equipment type.

## RESULTS

A rank ordering of the data was obtained for each judge on each of the stimulus pairs as a result of the pair-comparison task (see Appendix A). The rank-order data was then correlated between judges within each mission type and across both missions. The rank-order data for the stimulus pairs was also compared to the rank order of corresponding simulation outcomes, and the correlation matrices appear in Table 2.

As shown in the table, judges' ratings of attack data are in nearly perfect agreement. The judges' rankings of the loss data and the simulation outcomes calculated from the loss data also correlate nearly perfectly. Similar analyses of the covering force data also resulted in very high correlations, but not of the magnitude apparent in attack data. Finally, the magnitude of correlations are even lower when attack and covering force data are combined. Close inspection of judges' ratings show that J3 apparently had a general preference for covering force data, whereas J4 seemed to prefer attack data. This probably contributed to lower correlations across mission types as compared to those within mission types. The somewhat lower correlations in the covering force data with respect to attack data was apparently due to a factor not included in the simulated outcome scores for covering force data: whether or not the units remained combat effective. The judges explained that they rated loss data particularly low if they considered friendly forces to be combat ineffective at the conclusion of battle. Their criteria for combat effectiveness varied from about 50% to 30% of friendly forces surviving. The variation in judges' ratings of covering force battle outcomes appears to be partially attributable to differences in opinion as to what percent of friendly forces surviving constitutes being combat effective.

It appears that SMFRD and $\Delta$CR are particularly good at predicting judges' ratings of battle outcomes, where correlations range between .905 and 1.00, if mission type is held constant. It is, therefore, possible to combine the percentage of friendly forces surviving and percentage of OPFOR attrition into a single composite score, such as SMFRD or $\Delta$CR. These measures are, however, insensitive to the combat effectiveness factor. These data were used in the next phase of the current research to help develop a more comprehensive measure of battlefield performance, that include measures of covering force mission objectives.

## EXPERIMENT II

Least squares multiple regression was used to describe how the judges combined each of the covering force mission objectives in assigning mission accomplishment scores to the hypothetical battle outcome data. The covering force mission objectives, as defined in CATTS exercises, included: (1) attrit the enemy, (2) minimize friendly losses, (3) remain combat effective, (4) gather intelligence regarding the enemy's strengths and likely courses of action, (5) delay the enemy as far forward of the MBA as possible, (6) be prepared to conduct passage of lines, so as to take defensive positions in the MBA, and (7) avoid decisive engagements.

Table 2

Intercorrelations for Judges' Rankings and Simulation Outcomes

Attack

| | $J_1$ | $J_2$ | $J_3$ | $J_4$ | ΔCR | $C^2$ILL | SMFRD | RER |
|---|---|---|---|---|---|---|---|---|
| $J_1$ | -- | 1.00 | 1.00 | .98 | .98 | 1.00 | .98 | 1.00 |
| $J_2$ | | -- | 1.00 | .98 | .98 | 1.00 | .98 | 1.00 |
| $J_3$ | | | -- | .98 | .98 | 1.00 | .98 | 1.00 |
| $J_4$ | | | | -- | 1.00 | .98 | 1.00 | .98 |

Covering Force

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $J_1$ | -- | .91 | .93 | .88 | .91 | .93 | .93 | .86 |
| $J_2$ | | -- | .98 | .98 | 1.00 | .91 | .98 | .88 |
| $J_3$ | | | -- | .95 | .98 | .93 | 1.00 | .88 |
| $J_4$ | | | | -- | .98 | .88 | .95 | .81 |

Overall

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $J_1$ | -- | .91 | .60 | .65 | .90 | .85 | .95 | .65 |
| $J_2$ | | -- | .34 | .85 | 1.00 | .67 | .98 | .80 |
| $J_3$ | | | -- | -.13 | .29 | .86 | .46 | .89 |
| $J_4$ | | | | -- | .87 | .38 | .79 | .03 |

Experiment I demonstrated that objectives 1 and 2 could be combined into a single measure of relative loss between opposing forces, such as SMFRD and ΔCR. Judges indicated that SMFRD was easier to deal with conceptually, so SMFRD was used in this experiment to represent relative losses. Combat effectiveness (CE) was defined as ending battle above 50% of initial strength. Although somewhat arbitrary, this definition appears appropriate for the current context. Since it was not possible to obtain a good measure of decisive engagements, judges were instructed that if friendly forces were not combat effective at the conclusion of battle, to assume that they had been decisively engaged.

Quality of intelligence (INTEL) gathering was derived from OPFOR controllers' ratings on the following items, that were responded to by the battalion S2 near the conclusion of CATTS exercises: (1) estimate enemy strength, (2) estimate enemy location, (3) estimate enemy rate of advance, and (4) project the location of enemy attack on the MBA. The OPFOR controller, who maneuvered enemy elements, was in the best position to assess the responses of the above. Using a 5-point scale, the controller rated the S2 responses on a scale of 1 = very poor, to 5 = very good and the ratings were averaged across the four items to obtain a single value.

The degree to which BCGs were able to delay the OPFOR advance (OPFORD) was measured by calculating the average distance that OPFOR maneuver platoons were able to advance beyond the International Border (I.B). The relative location of friendly maneuver platoons with respect to the OPFOR (and the MBA) at the conclusion of battle was calculated by averaging the x-coordinate location of each friendly maneuver platoon. The distance between opposing forces (DBOF) was determined by subtracting this value from OPFORD.

The levels of the above variables used in this phase of the research appear in Appendix B. A SMFRD score of .000 is highest indicating that the percent of friendly and OPFOR surviving were equivalent at the conclusion of battle. Since the initial force levels favored the OPFOR by 3:1, this score indicated that friendly forces attrited the OPFOR at a rate of three times the losses incurred by friendly forces. A SMFRD score of -.500 could be obtained, for example, if at the conclusion of battle friendly forces were at 30% strength and the OPFOR was at 80% (-.500 = .300 -.800). It was expected that judges would treat SMFRD in a linear fashion where the least negative value represented the best performance.

Linear components of INTEL, OPFORD, and CE were also expected, where better intelligence gathering and less OPFOR penetration should be considered preferable. Being combat effective should be preferable to being noncombat effective.

The distance between opposing forces (DBOF) not only indicates how close friendly forces were to the enemy, but also how far they were from the MBA. Distance from the MBA can be calculated by subtracting the sum of OPFORD and DBOF from distance between the I.B. and the MBA. Since friendly forces are expected to be prepared to conduct passage of lines, it is not clear what values of DBOF are preferable.

## METHOD

### Subjects

The military experts who participated in Experiment I also served as judges in this phase of the research.

### Stimulus Materials

Stimulus materials were printed on 216 index cards which included the following hypothetical measures of performance in a three-hour battle (Appendix C): (1) a combined measure of percentage of friendly forces and OPFOR surviving battle (SMFRD), (2) whether or not friendly forces were combat effective at the conclusion of battle (CE), (3) the quality of intelligence gathering by friendly forces (INTEL), (4) the average distance of advance of the OPFOR platoons from the International Border (OPFORD) and (5) the average distance of friendly forces from the OPFOR (DBOF) at the conclusion of battle. These measures were selected to reflect the objectives of the covering force mission as stated to BCGs prior to the simulated battle.

### Procedures

Judges were presented with 216 stimulus cards containing the five hypothetical measures of battle performance, and were directed to rank-order the cards from worst to best in terms of mission accomplishment. At conclusion of this task, judges were instructed to assign a mission accomplishment score to each of the stimulus cards using a scale from 0 to 100. A value of 100 was to indicate perfect mission accomplishment, a value of 0 indicating total failure and a score of 50 reflecting average mission accomplishment in a covering force mission.

## RESULTS

Judges' ratings were subjected to linear regression analyses to derive regression equations that describe how each judge assigned mission accomplishment scores to the battle outcomes. Each equation was calculated using the University of California BMDP canned statistical runstream. BMD-9R-1981, "best set" multiple regression analysis was used to optimize accounted for variance with only linear components of main effects. Nonlinear components of main effects were included in equation for each judge until the individual regression equations accounted for ($R^2$) more than 95% of the variance in mission accomplishment scores.

These regression equations and a composite model appear in Table 3 along with the proportion of variance in mission accomplishment ratings that was accounted for by each equation. As indicated by the table, the proportion of variance accounted for exceeded 95% in all cases. For each judge, variance accounted for was due mostly to the linear, additive components of

the main effects, however, quadratic, cubic, and quartic components of the main effects were observed. SMFRD had a quadratic component for all judges. This trend was significant across judges and was included in the composite model. Also for Judge 2 there was a significant quadratic effect for DBOF. The effect was highly weighted (beta = -2.05) and in the opposite direction of the linear effect for that judge. Finally, it should be noted that J4's regression equation placed a high weight on the intelligence gathering factor, which also had cubic and quartic components.

The high correlations indicate that it is feasible to develop mathematical models which describe how judges combine several types of information in arriving at an overall assessment of battlefield performance. It is also apparent that the models are additive and void of important interaction effects. Before preceeding, however, it is of interest to examine the proportion of variance in mission accomplishment scores that was accounted for by only SMFRD and CE. As stated previously, SMFRD has been used in previous research as a candidate measure of mission accomplishment, and CE was found in Experiment I to be a measure of potential interest. Table 4 presents these data for each judge. These measures, individually and in combination, accounted for a large proportion of the variance in mission accomplishment scores for Judges 1, 2, and 3. Only a relatively small amount of additional variance is accounted for by the the linear regression equations (see Table 3).

The next phase of this research examined how well the regression models predicted judges' assessments on a similar task, which gave an indication of the predictive validity of the models.

## EXPERIMENT III

The final experiment was designed to determine if judges could apply their decision rules, as defined by the individual regression models, to a new set of battle outcome data taken from actual CATTS exercises. Interrater agreement on mission accomplishment scores was assessed by correlating the scores among judges. Also, individual models were combined into a composite model to predict average mission accomplishment scores.

### METHOD

#### Subjects

The military judges who participated in Experiments I and II served as subjects in this phase of the research.

#### Stimulus Materials

The stimulus materials, appearing in Appendix C, were 10 battle outcomes taken from actual CATTS exercises.

Table 3

Composite and Individual Models of Mission Accomplishment

$$y'(J1) = 66.83 + 75.164(SMFRD) + 26.264(CE) + 1.124(INTEL) - 1.803(OPFORD) - 0.041(DBOF) - 13.729(SMFRD)^2 \qquad R^2 = .99$$

$$y'(J2) = 11.294 + 18.296(SMFRD) + 40.785(CE) + 0.398(INTEL) - 2.131(OPFORD) + 21.494(DBOF) - 68.319(SMFRD)^2 - 2.051(DBOF)^2 \qquad R^2 = .96$$

$$y'(J3) = 54.503 + 103.497(SMFRD) + 19.797(CE) + 1.446(INTEL) - 0.319(OPFORD) - 0.775(DBOF) + 86.897(SMFRD)^2 \qquad R^2 = .98$$

$$y'(J4) = 269.545 + 34.311(SMFRD) + 12.845(CE) - 162.060(INTEL) - 0.814(OPFORD) - 0.495(DBOF) - 13.234(SMFRD)^2 + 16.206(INTEL)^3 - 2.239(INTEL)^4 \qquad R^2 = .99$$

$$y'(J1, J2, J3) = 44.213 + 65.652(SMFRD) + 28.948(CE) + 0.989(INTEL) - 1.418(OPFORD) + 6.893(DBOF) + 1.616(SMFRD)^2 - 0.684(DBOF)^2$$

Table 4

Variance in Mission Accomplishment Scores Accounted for by SMFRD and SMFRD and CE Combined

| | SMFRD | CE + SMFRD |
|---|---|---|
| J1 | .713 | .946 |
| J2 | .448 | .824 |
| J3 | .732 | .963 |
| J4 | .148 | .194 |

## Procedures

Judges were given their original 216 cards along with corresponding rank-order and mission accomplishment scores. By comparing the 10 stimulus cards to the original 216, judges were instructed to assign mission accomplishment scores to the battle outcomes derived from actual CATTS exercises.

## RESULTS

Of primary interest is the degree to which the linear regression models formulated in Experiment II predicted judges' ratings of the outcomes in Experiment III. The values of the components of mission accomplishment from actual CATTS exercises were substituted for unknown values in the regression models, multiplied by beta weights, and combined along with the value of the intercept, resulting in predicted mission accomplishment scores for each CATTS exercise for each judge. Predicted scores were correlated with actual mission accomplishment ratings for each judge. Correlations were squared to indicate the proportion of variance in actual ratings that were accounted for by predicted values. $R^2$ for Judges 1 through 4 were .937, .953, .964, and .266, respectively, indicating that the regression models accurately predicted actual mission accomplishment scores for three of four judges.

Although the individual regression models for each judge appear somewhat different, the judges' ratings of actual data were quite similar. The degree to which judges agreed on their actual mission accomplishment ratings was assessed by correlating these ratings among judges. The correlations among Judges 1, 2, and 3 were .94, .95, and .995, and for judge 4 compared to the others, r = .75, .75, and .76. It appears that judges 1, 2, and 3 were not only in agreement in their assessments of mission accomplishment, but also quite accurate in applying their models of mission accomplishment to new data. Conversely, J4 was not only deviant in his assessment of mission accomplishment, but also unable to reapply his decision rule,[1] as described by the regression formula, to actual data collected in BCG training exercises. For this reason, J4 was not included in the subsequent analysis.

Since it is desirable to develop measures of simulated battle performance for feedback and evaluation purposes, the individual regression models for Judges 1 through 3 were combined into a composite model (see Table 3). The composite of the individual models was derived by averaging beta weights from the individual models, resulting in a single regression equation. This

[1]Although Judge 4's strategy for assigning mission accomplishment scores to hypothetical data from Experiment II was comprehensive and consistent, it proved inappropriate for stimulus values not included in the hypothetical values. Values of mission objectives in actual CATTS data were typically different from those used in Experiment II.

was then recentered about the mean predicted mission accomplishment scores (based on the data from Experiment II) to determine the composite intercept. The measures collected in the CATTS exercises were then substituted for unknown values of the factors in the composite model to derive predicted mission accomplishment scores for each of the 10 exercises. Predicted scores were correlated with the average scores provided by the raters for each exercise. These values appear in Appendix D. There was a very high correlation between mission accomplishment scores predicted by the composite model and the average of the judges' ratings when judge 4 was included or omitted from the actual data ($r = .96$ and .98, respectively).

Finally, average mission accomplishment scores for the four judges were correlated with SMFRD across the ten CATTS exercises, resulting in $r = .81$. Since the composite model was better than SMFRD alone at predicting actual mission accomplishment scores, it can be concluded that the model represents a more comprehensive battlefield performance measure than SMFRD alone.

## GENERAL DISCUSSION

The multiple regression approach to describing how military judges combined various components of the covering force mission into a composite score of mission accomplishment appears highly successful. Regression equations including linear, quadratic, cubic, and quartic components of main effects only, accounted for over 95% of the variance in raters' assessments of overall mission accomplishment for all four judges. Although the procedure resulted in somewhat different regression formulas, the amount of inter-rater agreement in mission accomplishment ratings on the actual CATTS data was quite high. The individual regression models were very accurate in predicting judges' assessments of actual battle outcomes for three of four judges. And, even though the composite model of mission accomplishment was based on the individual models of only three judges, the composite accurately predicted average mission accomplishment scores of all four judges.

Mission accomplishment scores represent an advance over simple measures of relative losses such as SMFRD. Mission accomplishment scores not only take into account more of the mission objectives, but account for more variance in judges ratings than does SMFRD alone.

The utility of the composite model, that could predict mission accomplishment scores based upon determined levels of the component measures of covering force performance, would be to evaluate mission accomplishment of subsequent BCGs exercising in CATTS-like training exercises. Based on previous research on expert judgment, the composite model could be expected to more accurately assess mission accomplishment than the judges themselves. Relying on the model rather than the judges in future performance assessment could result in a more cost-effective assessment procedure.

Mission accomplishment scores and information as to how well BCGs performed on the components (mission objectives) of mission accomplishment could also be used as diagnostic feedback in after-action-reviews. Feedback to BCGs exercising in CATTS-like environments could focus on how and why the BCGs attained their scores on the mission objectives and overall mission accomplishment.

Although these results appear highly promising in terms of formulating composite measures of mission accomplishment in a $C^2$ training exercise, the results are based on a small sample of subjects, where one judge was unable to apply his decision rule to actual data. It is, therefore, desirable to replicate this research with a larger rater population to insure reliability of results. It is also desirable to extend the procedure to other mission types to ensure generalizability of the procedures.

## REFERENCES


Barber, H.F., and Kaplan, I.T. Battalion command group performance in simulated combat. ARI Technical Paper 353, March 1979. (AD A070 089)

Darlington, R.B. Multiple regression in psychological research and practice. Psychological Bulletin, 1968, 6, 161-182.

Dawes, R.M., and Corrigan, B. Linear models in decision making. Psychological Bulletin, 1974, 81, 95-106.

Dudycha, A.L., and Naylor, J.C. The effect of variations in the cue R-matrix upon the obtained policy equations of judges. Educational and Psychological Measurement, 1966, 26, 583-603.

Goldberg, L.R. Simple methods or simple processes? Some research on clinical judgments. American Psychologist, 1968, 23, 483-496.

Hammond, K.R., Hirsch, C.J., and Todd, F.J. Analyzing the components of clinical inference. Psychological Review, 1964, 71, 438-456.

Hirsch, C.J., Hammond, K.R., and Hirsch, J.L. Some methodological considerations in multiple-cue probability studies. Psychological Review, 1964, 71, 42-60.

Hoffman, P.J., Slovic, P., and Rorer, L.C. An analysis-of-variance model for the assessment of configural cue utilization in clinical judgment. Psychological Bulletin, 1968, 69, 338-349.

Kaplan, I.T., and Barber, H.F. Training battalion command groups in simulated combat: Identification and measurement of critical performance. ARI Technical Paper 376, June 1979. (AD A075 414)

Meehl, P.E. Clinical versus statistical prediction: A theoretical analysis and review of the literature. Minneapolis: University of Minnesota Press, 1954.

Meehl, P.E. Seer over sign: The first good example. Journal of Experimental Research in Personality, 1965, 1, 27-32.

Naylor, J.C., and Wherry, R.J., Sr. The use of simulated stimuli and the "JAN" technique to capture and cluster the policies of raters. Educational and Psychological Measurement, 1965, 25, 969-986.

Slovic, P., and Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 1971, 6, 649-744.

Thomas, G.S. Battle simulation outcomes as potential measures of BCG performance in CATTS exercises. ARI FLv FU Working Paper 83-1, March 1983.

Thomas, F.S., Barber, H.F., and Kaplan, I.T. The impact of CATTS system characteristics on selected measures of battalion command group performance. ARI Technical Report 609, U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A140 231)

## APPENDIX A

### Rank Order of Battle Outcomes by Judges

**Battle Outcomes**

| % Red Losses | % Friendly Surviving | Judge 1 | | Judge 2 | | Judge 3 | | Judge 4 | |
|---|---|---|---|---|---|---|---|---|---|
| | Attack | All Data | Attack Only | All Data | Attack Only | All Data | Attack Only | All Data | Attack Only |
| 6.6 | 72.0 | 7 | (4) | 6 | (4) | 11 | (4) | 4 | (4) |
| 6.8 | 68.4 | 8 | (5) | 8 | (5) | 12 | (5) | 5 | (5) |
| 6.9 | 60.2 | 12 | (6) | 11 | (6) | 13 | (6) | 9 | (7) |
| 6.9 | 77.3 | 4 | (2) | 3 | (2) | 9 | (2) | 2 | (2) |
| 3.4 | 58.5 | 14 | (8) | 13 | (8) | 16 | (8) | 11 | (8) |
| 3.1 | 65.4 | 13 | (7) | 10 | (7) | 15 | (7) | 6 | (6) |
| 7.0 | 73.9 | 5 | (3) | 4 | (3) | 10 | (3) | 3 | (3) |
| 6.6 | 87.9 | 3 | (1) | 1 | (1) | 8 | (1) | 1 | (1) |
| | Covering Force | | CF Only | | CF Only | | CF Only | | CF Only |
| 13.0 | 44.1 | 16 | (8) | 15 | (7) | 14 | (8) | 15 | (7) |
| 22.4 | 35.6 | 15 | (7) | 16 | (8) | 7 | (7) | 16 | (8) |
| 27.5 | 63.0 | 1 | (1) | 2 | (1) | 1 | (1) | 7 | (1) |
| 18.8 | 52.0 | 11 | (6) | 9 | (4) | 4 | (4) | 12 | (4) |
| 25.5 | 58.6 | 2 | (2) | 5 | (2) | 2 | (2) | 10 | (3) |
| 15.3 | 62.8 | 6 | (3) | 7 | (3) | 3 | (3) | 8 | (2) |
| 23.5 | 42.6 | 10 | (5) | 14 | (6) | 6 | (6) | 14 | (6) |
| 15.2 | 52.7 | 9 | (4) | 12 | (5) | 5 | (5) | 13 | (5) |

## APPENDIX B

### Levels of Battlefield Measures Used in Experiment II

| SMFRD | Combat Effective | Intelligence Gathering | Distance of OPFOR Advance | Distance Between Opposing Force |
|---|---|---|---|---|
| .000 | No = 0 | Poor = 2 | 10.0K | 2.0K |
| -.175 | Yes = 1 | Fair = 3 | 7.5K | 4.5K |
| -.340 | | Good = 4 | 5.0K | 6.5K |
| -.500 | | Very Good = 5 | | |

## APPENDIX C

### Stimulus Material for Experiment III

| No. | SMFRD | Combat Effective | Intel Gathering | OPFOR Advance Distance | Distance Between Opposing Forces |
|---|---|---|---|---|---|
| 1 | -.465 | No | 4.00 | 7.3K | 3.8K |
| 2 | -.043 | Yes | 5.00 | 7.8K | 6.7K |
| 3 | -.335 | No | 2.25 | 5.1K | 2.4K |
| 4 | -.175 | Yes | 3.50 | 9.6K | 4.1K |
| 5 | -.288 | Yes | 1.25 | 8.2K | 5.3K |
| 6 | -.049 | Yes | 4.00 | 7.7K | 3.1K |
| 7 | -.143 | Yes | 2.00 | 5.3K | 4.1K |
| 8 | -.334 | No | 2.50 | 8.2K | 1.2K |
| 9 | -.226 | No | 1.00 | 10.1K | .9K |
| 10 | -.018 | Yes | 4.00 | 6.3K | 4.3K |

## APPENDIX D

### Actual and Predicted Mission Accomplishment

| Unit | Day | Actuals (n=3) | Actuals (n=4) | Composite |
|---|---|---|---|---|
| 1 | 1 | 38 | 41 | 24 |
| 1 | 2 | 78 | 80 | 80 |
| 2 | 1 | 26 | 24 | 30 |
| 2 | 2 | 67 | 71 | 68 |
| 3 | 1 | 65 | 53 | 61 |
| 3 | 2 | 79 | 80 | 78 |
| 4 | 1 | 78 | 67 | 75 |
| 4 | 2 | 25 | 25 | 21 |
| 5 | 1 | 27 | 24 | 22 |
| 5 | 2 | 79 | 80 | 84 |