AFHRL-TR-84-39

(12)

AIR FORCE HUMAN RESOURCES LABORATORY

AD-A149 285

CLOSE AIR SUPPORT MISSION: DEVELOPMENT OF A UNITARY MEASURE OF PILOT PERFORMANCE

By

Gary S. Thomas

OPERATIONS TRAINING DIVISION
Williams Air Force Base, Arizona 85240-6457

November 1984
Interim Report for Period February 1984 - July 1984

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5000

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

MILTON E. WOOD, Technical Director
Operations Training Division

ANTHONY F. BRONZO, JR., Colonel, USAF
Commander

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

AD-A149285

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFHRL-TR-84-39	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Operations Training Division Air Force Human Resources Laboratory 6b. OFFICE SYMBOL (If applicable) AFHRL/OT	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State and ZIP Code) Williams Air Force Base, Arizona 85240-6457	7b. ADDRESS (City, State and ZIP Code)
8a. NAME OF FUNDING SPONSORING ORGANIZATION Air Force Human Resources Laboratory 8b. OFFICE SYMBOL (If applicable) HQ AFHRL	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State and ZIP Code) Brooks Air Force Base, Texas 78235-5000	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO. 62205F; PROJECT NO. 1123; TASK NO. 11; WORK UNIT NO. 27

11. TITLE (Include Security Classification)
Close Air Support Mission: Development of a Unitary Measure of Pilot Performance

12. PERSONAL AUTHOR(S)
Thomas, G.S.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Interim	FROM Feb 84 TO Jul 84	November 1984	22

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB. GR.		
05	08		Advanced Simulator for Pilot Training	close air support
05	09		air to surface	combat simulation
			A-10	expert judgment

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

This effort demonstrated the feasibility of combining components of the Close Air Support (CAS) mission in a valid composite measure of pilot performance. Combat-ready, A-10 aircraft pilots judged overall mission performance from outcomes typical of those obtained in the Advanced Simulator for Pilot Training (ASPT) based on pilot survival and the number and type of targets destroyed. A linear regression analysis of the judgments resulted in a mathematical formula that assigned differential values to the mission components. The formula was cross-validated using a separate group of A-10 pilots who judged performance for a different set of CAS mission outcomes. To test the sensitivity of the scoring algorithm to CAS training, data from a previous study on pilots trained in ASPT were reanalyzed using the formula. CAS performance, as calculated by the algorithm, improved significantly across training trials. This procedure for developing a unitary performance metric may be useful in other training research and development when a composite measure is desirable for evaluation purpose. Originator-supplied keywords include: Air to surface, Combat simulation, and Expert judgment.

20. DISTRIBUTION AVAILABILITY OF ABSTRACT UNCLASSIFIED UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION
22a. NAME OF RESPONSIBLE INDIVIDUAL Nancy A. Perrigo Chief, STINFO Office	22b. TELEPHONE NUMBER (Including Area Code) (512) 536-3877 22c. OFFICE SYMBOL AFHRL/TSR

DD FORM 1473, 83 APR

EDITION OF 1 JAN 73 IS OBSOLETE.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

Item 18 (*Continued*)

performance measurement performance prediction pilot training simulation training

SUMMARY

The objective was to develop and validate a composite measure of Close Air Support (CAS) mission performance in support of on-goirg and anticipated training research and development (R&D) using the Advanced Simulator for Pilot Training (ASPT). A linear regression analysis was used to describe how mission-ready, A-10 aircraft pilots rank-ordered hypothetical CAS mission outcomes typical of those obtained in ASPT. CAS performance ratings predicted by the regression model were then compared to actual rankings assigned to a larger set of CAS outcomes by a different group of A-10 pilots. On the average, the model accounted for 93% of the variance in pilots' rankings. To test the sensitivity of the regression model to CAS training obtained by pilots exposed to ASPT, the data collected in a previous study were reanalyzed using the model. CAS performance as calculated by the model was found to improve significantly across training trials.

The linear regression approach to describing how expert judges determine overall performance based on several components of the CAS mission appears highly successful. The regression model of CAS performance has good predictive qualities and is sensitive to training provided by ASPT. Finally, the procedures used in this effort may be appropriate for use in other measurement R&D where a unitary measure of performance is desirable.

A-1

PREFACE

This effort represents a portion of the Air Force Human Resource Laboratory research and development (R&D) program for Technical Planning Objective 3, the thrust of which is Aircrew Training. The general objective of this thrust is to identify and demonstrate cost-effective training strategies and training equipment capabilities for use in developing and maintaining the combat effectiveness of aircrew members. More specifically, the effort was part of the R&D program conducted under the Air Combat Training Research subthrust, which has as its goal to provide a technical base for training high-level and quickly perishable skills in simulated combat environments. Work Unit 11231127, Electronic Warfare Training Effectiveness, addressed a portion of this subthrust; namely, the use of a simulated combat scenario for close air support training.

The author expresses appreciation to Dr. David Hubbard for his analysis of the data, to Dr. Lowell Schipper and Dr. Wayne Waag for their critical review, and to Dr. Elizabeth Martin for her assistance in designing the research. Thanks also go to the A-10 pilots whose cooperation made the study possible.

TABLE OF CONTENTS

LIST OF TABLES

CLOSE AIR SUPPORT MISSION: DEVELOPMENT OF A UNITARY MEASURE OF PILOT PERFORMANCE

I. INTRODUCTION

A prerequisite to conducting close air support (CAS) training research and development (R&D) is the ability to measure relevant pilot performance in the CAS mission. Developing a single summary measure based on the components of mission success has several advantages in its application to training R&D. Such a measure could (a) give a general indication of pilot proficiency prior to training exercises, (b) be used to scale difficulty of candidate training exercise scenarios, (c) be used to evaluate alternate training procedures, and (d) provide a general measure of performance for feedback to trainees.

Recent advances in training technology and computer-generated visual displays make combat training in flight simulators an option to at least supplement training provided by range exercises. Kellogg, Prather, and Castore (1981) demonstrated that the Advanced Simuator for Pilot Training (ASPT), configured as an A-10 aircraft cockpit, could be used for training penetration, attack, and egress against a tank defended by anti-aircraft artillery (AAA) and surface-to-air missiles (SAMs). Hughes, Brooks, Graham, Sheen, and Dickens (1982) reported that A-10 pilots who received CAS mission training in ASPT prior to "Red Flag" combat exercises were more likely to survive simulated combat on the range than were pilots who did not receive pretraining in ASPT. These experiments used separate measures of offensive and defensive performance.

In order to determine further the potential of ASPT to supplement CAS training available in range exercises, several R&D issues should be addressed. For example, it is desirable to provide the optimal training exercise to the pilot based on his level of expertise. This requires a pretraining metric that is sensitive to the strengths and weaknesses in a pilot's performance, so that a training experience can be provided that exercises those skills which need the most practice. Procedures for providing performance feedback are also necessary to maximize training benefit from the simulator.

Procedures for determining how multiple sources of information are combined into overall assessments can be found in the decision-making literature (see Slovic & Lichtenstein, 1971, for a review). In particular, least-squares linear regression has been used in a variety of settings to describe how expert judges arrive at composite assessments of performance. 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 do 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. A linear model that can be constructed for all judges is a better predictor of judges' ratings than is any single model obtained from individual judges.

Thomas and Cochlin (1983) used a regression model to describe how expert Army judges combined measures of several components of a defensive maneuver into a single measure of overall mission accomplishment. Hypothetical battle outcomes were used to develop the models of individual judges' decision rules. Validity of the predictive models was determined by comparing each judge's ratings of actual battle outcomes, generated by battalion command groups conducting covering force missions in a computer-driven battle simulation, to ratings of mission accomplishment predicted by the models. The individual models accounted for at least 94% of the variance in three of four judges' ratings of actual outcomes. When the individual models were combined into a single model, that model accounted for 92% of the variance in the average of the judges' ratings of actual outcomes.

The purpose of the current effort was to apply the regression approach to the development of a valid composite measure of CAS mission performance reflecting the primary objectives of the A-10 aircraft mission: to destroy enemy tanks while avoiding/evading threats.

II. EXPERIMENT I

This first experiment was concerned with developing a simple scoring formula that could assess overall CAS performance by taking into account the relevant components of the mission. The scoring metric would assign a number of points for each component outcome, such as surviving the mission or destroying tanks or threats. The summation of points would result in an overall performance score for each mission. To determine the appropriate number of points to assign to each component outcome, expert judgments were elicited from mission-ready, A-10 pilots. The judgments were then analyzed by linear regression analysis to derive a performance metric.

Method

Subjects

Eight, mission-ready, A-10 pilots served as expert judges. Pilots had flown from 200 to 400 hours in the A-10 aircraft, and all but one had experience in at least one range exercise involving CAS missions. All pilots also had 2 hours of experience in ASPT CAS missions prior to participation in this experiment.

Stimulus Materials

Stimulus materials were 117 cards (3 by 5 inches), each listing a hypothetical A-10 CAS mission outcome. Variables included on the cards were number of tanks and threats destroyed, time survived in the environment, whether the pilot survived his mission (0 or 1), and whether a CP was destroyed (0 or 1). The number of tanks and threats destroyed ranged from 0 to 5, and the time-survived values were 30 seconds, 100 seconds, 170 seconds, or 240 seconds, where only the 240-second condition implied that the pilot survived the mission. Total targets destroyed (combinations of tanks, threats, and the CP) ranged from 0 to 5. Stimulus cards, with time-survived values of 100 seconds, 170 seconds, and 240 seconds, included all possible combinations of 0 to 5 targets destroyed. For purposes of realism, the 30-second condition included all possible combinations of 0 to 2 targets destroyed.

Procedure

Judges were instructed that the purpose of the experiment was to develop a unitary measure of CAS mission accomplishment based on the variables included on the stimulus cards. Judges ranked the 117 combat outcomes from best to worst in terms of how well the hypothetical CAS missions were accomplished.

Results

Comparison of Judges' Responses

Spearman rho correlations were calculated comparing the rank-orders of combat outcomes obtained from the eight judges. As indicated by correlations in Table 1, there was relatively high agreement among judges as to their rankings of combat outcomes. Correlations ranged from .68 to .99, with a median rho of .90.

Table 1. **Intercorrelations Among Judges' Rankings of CAS Mission Outcomes**

Judge	A	B	C	D	E	F	G	H
A	-	.860	.684	.903	.819	.896	.795	.802
B	-	-	.923	.961	.989	.917	.980	.901
C	-	-	-	.881	.919	.851	.953	.807
D	-	-	-	-	.922	.984	.947	.789
E	-	-	-	-	-	.864	.967	.927
F	-	-	-	-	-	-	.912	.716
G	-	-	-	-	-	-	-	.854
H	-	-	-	-	-	-	-	-

Even though judges tended to agree in their assessments of the combat outcomes, the data were also analyzed by a mixed-model covariance analysis (using the unique sums of squares approach) to specify further where judges agreed and disagreed on the importance they attached to the variables used in the stimulus materials. It was determined that judges disagreed on the relative importance of destroying tanks, threats, and the CP. The analysis indicated significant subjects-by-tanks, subjects-by-threats, and subjects-by-CP interactions ($P < .0001$, source table appears in Appendix A).

On the other hand, judges did not disagree in their treatment of the mission survival or time-survived variables. All agreed, for example, that it was better to survive and not destroy any targets, than it was to destroy numerous targets and not survive the mission. This effect was exemplified by a nonsignificant subjects-by-mission-survival interaction ($F_{7, 888} = .97$). Finally, the time survived in the environment variable was *not* treated differentially by judges. It was generally perceived that, all else being equal, it was better to survive longer, as indicated by a nonsignificant subjects-by-time interaction ($F_{7, 888} = 1.06$). In summary, judges did not disagree on the importance they associated with mission survival and time survived, but some judges disagreed on the relative importance of destroying tanks versus threats versus the CP.

Impact of Mission Variables on Judges' Rankings

The covariance analysis (Appendix A) also demonstrated that all variables had a significant impact on judges' assessments of CAS missions outcomes ($P < .0001$). That is to say, mission survival, more time survived in the environment, destroying the CP, and destroying more versus fewer tanks and threats resulted in higher assessments of mission success by the judges.

Regression Analyses

Judges' rankings of the CAS mission outcomes were also subjected to linear regression analysis to derive regression equations that describe how each rank-ordered the CAS outcomes. An equation for each judge was calculated using a statistical package (BMDP-9R) to optimize accounted-for-variance with a "best fit," step-wise regression analysis including only linear components of main effects. The regression equations accounted for 95 to 97 percent of the variance in individual judges' rankings of mission outcomes. Individual regression equations appear in Appendix B.

The differences in b weights associated with the CP, threats, and tanks variables again point to the differences in relative importance of these targets in the opinion of the judges. However, it is noteworthy that the equations accounted for such high proportions of the variance in the data when considering only linear components of main effects. Apparently, the regression approach is highly satisfactory in describing how judges determine what is important in a successful CAS mission.

The primary purpose for conducting this experiment was to derive a measure of CAS mission performance based on the expert opinions of experienced A-10 pilots. To achieve that purpose, the data from all judges were combined and again analyzed by least-squares linear regression. The resulting regression equation was:

$$\text{Performance Scores} = -.37 + 51.6\,(\text{mission survival}) + 18.3\,(\text{CP}) + 10.8\,(\text{tanks}) + 9.7\,(\text{threats}) + .04\,(\text{seconds survived})$$

The equation accounted for 88% of the variance in the raw data, and 98% of the variance in the average rankings of judges by considering only linear components of main effects.

Mission accomplishment scores for each of the original 117 combat outcomes were calculated by substituting frequency values for each variable into the composite equation and multiplying by the appropriate b weights. The resultant scores were then rank-ordered and correlated with the rank-orders obtained from the judges to determine the degree to which the composite model predicted judges' responses. As indicated in Table 2, correlations between judges' rank-orders of mission outcomes and those derived from the composite model are quite high. The very high correlations demonstrate that the composite model does very well at representing judges' assessments of CAS mission performance.

Table 2. **Comparison of Judges' Assessments to those Predicted by the Composite Model**

Judges	A	B	C	D	E	F	G	H
Correlations	.889	.996	.903	.965	.983	.928	.972	.898

Discussion

The least-squares regression model for CAS mission performance developed in this effort appears to be successful in describing how expert judges (trained A-10 pilots) combined multiple sources of combat information to determine overall mission performance. Even though there was not total agreement among the judges as to the relative importance of the components of CAS mission performance, the high correlations between predictions of the composite model and the assessments of the judges indicate that the model successfully represents judges' decision rules.

Although the results are encouraging in terms of mathematically representing decision rules applied to combat outcomes by expert judges, the model is based on a relatively small number of judges and a restricted set of hypothetical combat outcomes. To demonstrate that the model is a valid measure of CAS mission performance, it was desirable to cross-validate these results with a larger sample of combat outcomes and with additional judges.

III. EXPERIMENT II

The second experiment was an attempt to validate the CAS performance model developed in Experiment I using different A-10 pilots as subjects and a larger set of stimulus materials. The model was simplified in three respects: (a) the constant term (-.37) was deleted since it served no practical purpose, (b) the time survived in the environment component was deleted since the variable contributed only .002 to the variance accounted for by the model, and (c) the beta weight for the mission survival variable was increased to 150 to fit the new stimulus set that included larger values for tanks and threats. For example, in Experiment I all judges considered the outcome where the mission was survived (240 seconds) but no targets were destroyed to be superior to the outcome where four tanks and one threat were destroyed but the mission was not survived (time survived = 170 seconds). The model would predict the following performance scores for the two outcomes: 61.2 and 60.0, respectively. To maintain the same relationship in the stimulus set used in Experiment II, mission survival was assigned a value of 150, so that surviving the mission with no targets destroyed would result in a score of 150. Destroying 12 tanks and two threats but not surviving the mission would result in a lesser score: 149.7. The scale of possible scores would then range from 0 to 300, with a score of 150 as the mid-point.

Method

Subjects

Ten mission-ready, A-10 pilots served as expert judges. These pilots had from 180 to 1000 hours of experience in the A-10 aircraft, and all but one had at least one range experience involving CAS missions. All pilots had at least 2 hours of CAS training in ASPT prior to participation in this experiment.

Stimulus Materials

The stimulus materials were hypothetical CAS missions outcomes listed on 360 cards (3 by 5 inches). The mission variables were the number of tanks and threats destroyed, whether the mission was survived, and whether a CP was destroyed. Values for the latter two variables were either yes or no (1 or 0). The number of tanks destroyed varied from 0 to 12, and the number of threats (AAA or SAMs) ranged from 0 to 8. The maximum number of targets destroyed was limited to 14. The stimulus set included all possible combinations of the variables, given the limits described above.

The stimulus cards were sorted into sequential order based on the following criteria: (a) whether the mission was survived, (b) total number of targets destroyed, (c) number of tanks destroyed, (d) number of threats destroyed, and (e) whether the CP was destroyed. Five representative samples of the stimulus set were obtained by selecting out every fifth card without replacement, every fourth card without replacement, etc. Two combat outcomes of particular interest were then added to each sample set: an outcome where no targets were destroyed, but the mission was survived; and an outcome where no targets were destroyed and the mission was not survived. Each of the five sample sets was presented to 2 of the 10 judges so that each judge viewed 72 outcomes.

Procedure

The judges were instructed that the purpose of the experiment was to develop a unitary measure of CAS mission performance based on the variables included on the stimulus cards. Judges ranked the combat outcomes from best to worst in terms of how well the CAS missions were accomplished.

Results

A predicted rank-order of CAS outcomes was obtained based on the revised performance model developed in Experiment I for each of the five sample sets. Predicted rank-orders were then correlated with the rank-orders provided by the 10 judges. The correlations that appear in Table 3 are all quite high, indicating that the model does well at predicting judges' rankings of a new set of combat outcomes. The correlations between actual and predicted ratings range from .889 to .987 with a median of .964. The average amount of variance accounted for by the model was 93 percent.

Table 3. **Correlations Between Predicted and Actual Rank-Orders of CAS Mission Outcomes**

Judges	1	2	3	4	5	6	7	8	9	10
Correlations	.957	.987	.968	.949	.965	.964	.965	.964	.889	.961

Even though the model developed in Experiment I was based on a restricted set of data, the model very accurately predicted ratings made by expert judges evaluating the full range of outcomes expected in the simulated CAS mission. To describe judges' ratings in this experiment, all the data were subjected to a linear regression analysis. The resulting regression equation was as follows:

Performance Scores = .922 + 35.91 (mission survival) + 4.11 (CP) + 2.83 (tanks) + 1.34 (threats)

This model accounted for 94 percent of the variance in the data (R^2 = .94) by considering only linear components of main effects.

To facilitate further comparison of this model with that derived in Experiment I, all b weights were multiplied by 4.177. This constant was obtained by dividing the coefficient for the mission survival variable in the above equation (35.91) into the coefficient for the same variable in the model derived in Experiment I (150). The resulting equation along with the model derived in Experiment I appears in Table 4. As shown in the table, the two equations (ignoring the constant factors) are quite similar.

Table 4. *Regression Equations Derived From Experiments I and II*

Experiment I	Y = 150 (survival) + 18.27 (CP) + 10.85 (tanks) + 9.75 (threats)
Experiment II	Y = 150 (survival) + 17.15 (CP) + 11.82 (tanks) + 5.60 (threats)

Discussion

It appears that the performance model developed in Experiment I, even though it was based on a small sample of potential CAS outcomes, does well at predicting subsequent judges' rankings of a

larger set of combat outcomes. Correlations between actual and predicted rankings averaged .964. The regression approach to describing how judges assigned ranks to the outcomes in Experiment II was also successful. The regression equation accounted for 94 percent of the variance in the data. Finally, the regression equation derived in Experiment II, when multiplied by an appropriate constant, closely approximates the equation developed in Experiment I. This indicates that very similar predictor equations can be obtained by two different procedures. In Experiment I, all subjects observed the same relatively small set of stimuli, whereas in Experiment II, the judges observed a different but representative sample of all possible CAS outcomes.

IV. EXPERIMENT III

The mathematical model of CAS performance developed in Experiment I and validated in Experiment II appears to do well at describing how expert judges assess hypothetical CAS performance. For the model to be useful as a research tool or an evaluation metric in ASPT, the model should be responsive to experimental manipulations used in ASPT research.

ASPT has demonstrated potential for training A-10 pilots in the CAS mission. In particular, Hughes et al. (1982) reported that A-10 pilots who pretrained in ASPT and who were allowed to use chaff at "Red Flag" combat exercises survived more missions at "Red Flag" than did comparable pilots who did not receive ASPT pretraining. The pretrained pilots obtained about 1 hour of training in a CAS mission and about 1 hour of training in an interdiction mission where the primary objective was a CP. About half the pilots received CAS training prior to interdiction training, and the other half received interdiction training first.

To be useful, the CAS performance model should be sensitive to training such as that provided by Hughes et al. (1982) in ASPT. That is to say, performance scores calculated with the model should increase across training trials. The Hughes, et al. (1982) data were reanalyzed to test this notion; that is, performance scores were calculated for each A-10 pilot on each CAS training trial conducted by Hughes, et al. (1982) and performance was tracked across trials.

Method

Of the 17 pilots trained by Hughes, et al. (1982) data were retrievable from magnetic tapes for 15 pilots. Nine of these pilots received interdiction mission training prior to CAS training, whereas the other six obtained CAS training first.

The number of tanks and threats destroyed and whether pilots survived the missions or destroyed the CP were noted for each pilot on each mission flown. These data were substituted into the mathematical model of performance to calculate a performance score for each mission. Performance scores were averaged across the first half and the last half of the CAS training trials for each A-10 pilot.

Results

Average performance scores on first-half and second-half trials for pilots receiving CAS training first or interdiction training first are presented in Table 5. As shown in the figure, performance was better on second-half trials than on first-half trials. Also, pilots who received interdiction training prior to CAS training generally performed better than did pilots who received CAS training before interdiction training. To test these effects statistically, data were cast into a 2 x 2 mixed Analysis of Variance (ANOVA) with repeated measures on the training trial factor. As demonstrated by the analysis (source table appears in Appendix C), performance

was significantly better on second-half trials ($F_{1,\ 13} = 15.64$, $P < .005$), indicating improved performance as a result of training. The performance of the pilots who received interdiction training first was also significantly better than that for pilots receiving CAS training first ($F_{1,\ 13} = 20.55$, $P < .001$). Apparently, interdiction training had a facilitating effect on subsequent CAS performance. Finally, no interaction between the two main effects was observed ($F_{1,\ 13} = .005$).

Table 5. Mean CAS Performance Scores for First-Half and Last-Half Training Trials Conducted Before or After Interdiction Training

CAS	First-Half Trials	Last Half Trials
Before	17.9	44.6
After	49.6	75.9

Discussion

The mathematical model of CAS mission performance appears sensitive to CAS training provided by ASPT. Pilot performance, as assessed by the model, was significantly better in later training trials than in earlier trials. Also of interest is the fact that performance improves at about the same rate for both groups of pilots regardless of whether they received interdiction training prior to CAS training. The apparent facilitation in CAS performance as a result of prior interdiction training was substantial. Perhaps many of the same pilot skills and behaviors required in the CAS mission can be practiced in the interdiction mission.

IV. A SIMPLIFIED MODEL OF CAS PERFORMANCE

The following discussion concerns the development of a simplified model of CAS mission performance that may be useful for CAS training R&D, CAS performance evaluation, or as feedback to pilots receiving CAS training in ASPT. Data from Experiment I were reanalyzed using regression analysis and excluding the time-survived factor. When the coefficient for the mission survival variable was raised to 150, the model accounted for 87 percent of the variance in the data. This equation was combined with the equation developed in Experiment II (see Table 5). Corresponding b weights were simply averaged across the two formulas and were rounded to the nearest whole number, resulting in the following equation: Y = 150 (survival) + 18 (CP) + 11 (tanks) + 8 (threats).

To test the predictive quality of the simplified model, performance scores were calculated for each of the mission outcomes presented to the judges in each experiment. Scores were calculated by multiplying appropriate b weights by corresponding frequencies for targets destroyed, and then summing the products with 150 or 0 depending upon whether the mission was survived.

Performance scores were then correlated with rankings assigned to corresponding mission outcomes for each judge in the separate experiments. The multiple correlation and variance accounted for in Experiment I were 88 and 78 percent, respectively. As stated previously, hypothetical outcomes used in Experiment I represented only the lower one-fifth of possible outcomes when the mission was and was not survived. These restricted ranges of values may have reduced the correlation between actual and predicted ratings.

When the full range of potential outcomes was presented to judges, as was the case in Experiment II, there was a higher degree of relationship between performance scores and judges' rankings. In Experiment II, the multiple correlation was .97 and the accounted-for variance was

93%. The data analyses indicated, therefore, that even a simplified model which can easily calculate performance scores accurately represents judges' assessments of CAS mission outcomes. The simplified model results in a performance scale of 0 to 300 where the mid-point, a score of 150, can be achieved by surviving a mission while destroying no targets.

VI. CONCLUSIONS

The multiple regression approach to describing how Air Force judges combined various components of the CAS mission in assessing overall mission performance appears highly successful. The composite model of mission performance accounted for 98% of the variance in the average of judges' rankings in Experiment I when considering only linear components of main effects. Although the approach resulted in somewhat different regression formulas for individual judges, the amount of inter-rater agreement in CAS performance rankings was quite high. The composite model also did quite well at predicting the responses of a different set of judges assessing a different set of CAS outcomes. The median correlation between predicted and actual rankings made by the judges was rho = .964. It was also demonstrated that a simplified composite model could be developed to predict data from both sets of judges.

The model appears sensitive to the training resulting from repeated exposures to the CAS mission in ASPT. In Experiment III, pilot performance improved significantly as a function of CAS training. In addition, the pilots benefited from pretraining in the interdiction mission. Those who received pretraining in interdiction performed significantly better in the CAS mission than those who did not.

Potential Applications

The simplified model should be useful as a general metric to evaluate performance of future A-10 pilots conducting CAS missions in ASPT. The model may be useful in discriminating the better performing pilots, so that training exercises could be tailored to the proficiency level of the individual pilot. The model may also be of value in discriminating difficult exercises from relatively easy ones, because better performance would be expected on the less challenging exercises. More difficult exercises could then be provided to pilots who are initially "higher on the learning curve."

It has been demonstrated that the model is sensitive to training, therefore it should also be sensitive to different CAS training conditions used in ASPT. Better training techniques should result in better performance scores calculated by the model. These performance scores could be presented to pilots at the conclusion of each CAS training trial. This type of feedback may be useful in shaping desired pilot performance in ASPT. Or, individual pilot performance when referenced against average pilot performance may serve as an incentive for individuals to perform better in subsequent training trials.

One major limitation of the performance model is that it considers only battle outcomes and not those pilot behaviors necessary to effect these outcomes. Future R&D could use performance scores generated by the model as criterion scores to be compared against specific pilot behaviors. In other words, pilot behaviors that result in good criterion performance could be isolated from those that do not. Future training could focus on exercising and shaping those behaviors to maximize the training potential of ASPT.

Finally, the techniques of describing expert judges' assessments of mission performance using linear regression could be applied to other types of missions, such as interdiction or air-to-air engagements. The procedure appears successful in capturing the decisions used by experts in assigning relative weights to various components of mission performance.

REFERENCES

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

Hughes, R.G., Brooks, R., Graham, D., Sheen, R., & Dickens, T. (1982). Tactical ground attack: On the transfer of training from flight simulator to operational Red Flag range exercise. In Proceedings of the Fourth Interservice/Industry Training Equipment Conference and Exhibition, Orlando, Florida.

Kellogg, R.S., Prather, D.C., & Castore, C.H. (1981). Simulated A-10 combat environment. In Proceedings of the 1981 Image Generation/Display Conference II (pp. 36-44). Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory.

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

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

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

Thomas, G.S., & Cochlin, T. G. (1983), Performance appraisal in a battle simulation. In Proceedings of the Human Factors Society 27th Annual Meeting, (pp. 901-905). Norfolk, Virginia.

APPENDIX A

Covariance Analysis for Experiment I

Source of Variance	Sum of Squares	df	Mean Squares	F	P
Subject	2,555	7	365	7.4	.000
Time	1,901	1	1,901	36.2	.001
Tanks	159,363	1	159,363	87.5	.000
Threats	128,738	1	128,738	39.2	.000
Command Post	71,876	1	71,876	15.3	.006
Survival	176,078	1	176,078	3,569.5	.000
Subj x Time	367	7	52	1.1	.385
Subj x Tanks	12,742	7	1,820	36.9	.000
Subj x Threats	23,014	7	3,288	66.6	.000
Subj x Cmd Post	32,846	7	4,692	95.1	.000
Subj x Survival	335	7	48	1.0	.452
Within + Residual	43,814	888	49		

APPENDIX B

Regression Equations for Individual Judges

A	$Y = 9.17 + 54.02$ (survival) $+ 36.47$ (CP) $+ 7.72$ (tanks) $+ 1.92$ (threats) $+ .03$ (time) $R^2 = .965$
B	$Y = -6.01 + 51.58$ (survival) $+ 18.20$ (CP) $+ 13.45$ (tanks) $+ 11.41$ (threats) $+ .03$ (time) $R^2 = .965$
C	$Y = .96 + 50.72$ (survival) $+ 2.16$ (CP) $+ 9.46$ (tanks) $+ 14.25$ (threats) $+ .04$ (time) $R^2 = .946$
D	$Y = -2.24 + 54.28$ (survival) $+ 29.54$ (CP) $+ 9.25$ (tanks) $+ 11.46$ (threats) $+ .01$ (time) $R^2 = .960$
E	$Y = -5.16 + 49.95$ (survival) $+ 12.08$ (CP) $+ 14.77$ (tanks) $+ 10.99$ (threats) $+ .04$ (time) $R^2 = .960$
F	$Y = -.48 + 52.52$ (survival) $+ 32.10$ (CP) $+ 6.05$ (tanks) $+ 11.10$ (threats) $+ .03$ (time) $R^2 = .967$
G	$Y = -8.17 + 46.88$ (survival) $+ 13.07$ (C) $+ 11.26$ (tanks) $+ 13.00$ (threats) $+ .08$ (time) $R^2 = .952$
H	$Y = -9.27 + 52.49$ (survival) $+ 2.57$ (CP) $+ 14.80$ (tanks) $+ 3.86$ (threats) $+ .03$ (time) $R^2 = .955$

APPENDIX C

ANOVA Table For Red Flag Data

Source of Variance	Sum of Squares	df	Mean Squares	F	P
Total	37,933.77	29			
Between	28,046.13	14			
CAS 1st vs 2nd	17,177.62	1	17,177.62	20.55	.001
Between Error	10,868.51	13	836.04		
Within	9,887.64	15			
Training Trials	5,400.21	1	5,400.21	15.64	.005
Training x CAS	.13	1	.13	.005	
Within Error	4,487.30	13	345.18		