June 1990

PREDICTING F-14 AIR COMBAT MANEUVERING (ACM) PERFORMANCE USING AN AUTOMATED BATTERY OF COGNITIVE/PSYCHOMOTOR TESTS

R.N. SHULL AND G.R. GRIFFIN

Naval Aerospace Medical Research Laboratory
Naval Air Station
Pensacola, Florida 32508-5700

Reviewed and approved 1 June 1990

J. A. Brady

J. A. BRADY, CAPT, MSC USN
Commanding Officer

The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government.

Trade names of materials and/or products of commercial or non-government organizations are cited as needed for precision. These citations do not constitute official endorsement or approval of the use of such commercial materials and/or products.

Volunteer subjects were recruited, evaluated, and employed in accordance with the procedures specified in Department of Defense Directive 3216.2 and Secretary of the Navy Instruction 3900.39 series. These instructions are based upon voluntary informed consent and meet or exceed the provisions of prevailing national and international guidelines.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
June 1990

3. REPORT TYPE AND DATES COVERED
Interim Mar 88 - Mar 89

4. TITLE AND SUBTITLE
Predicting F-14 Air Combat Maneuvering (ACM) Performance Using an Automated Battery of Cognitive/Psychomotor Tests

5. FUNDING NUMBERS
63706N
M0096.001 0151

6. AUTHOR(S)
R.N. Shull and G.R. Griffin

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Aerospace Medical Research Laboratory
Bldg. 1953, Naval Air Station
Pensacola, FL 32508-5700

8. PERFORMING ORGANIZATION REPORT NUMBER
NAMRL-1356

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Medical Research and Development Command
National Naval Medical Center, Bldg. 1
Bethesda, MD 20814-5044

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

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

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
Some studies have suggested the possibility of predicting operational performance in fleet aviation environments. The current report concerns the use of an automated performance-based test battery, involving cognitive and psychomotor functioning, to predict the operational performance of fighter pilots. A group of jet pilots completing Air Combat Maneuvering (ACM) training in the F-14 were tested on this battery. The few significant correlations found between the test measures and ACM performance measures were illogically patterned and of insufficient quantity or strength to establish that such a battery would reliably predict ACM performance. This could have been due to the homogeneous nature of the subject group in terms of pilot skills and abilities. Given these results, this particular test battery would probably not be useful in the prediction of flight performance at such a late stage of training as ACM, but it might predict flight performance in earlier training.

14. SUBJECT TERMS
Pilot Selection, Cognition, Psychomotor
Dichotic Listening, Air Combat Maneuvering, F-14, Jets

15. NUMBER OF PAGES
9

16. PRICE CODE

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

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

SUMMARY PAGE

THE PROBLEM

Several studies have suggested the possibility of predicting operational performance in fleet aviation environments. Research is currently being conducted to develop reliable predictor tests that might aid in decisions concerning aircrew selection, training pipeline assignment, and posttraining aircraft assignment. The current approach involves using an automated test battery, which measures various aspects of cognitive and psychomotor functioning, to predict the operational performance of fighter pilots beyond advanced undergraduate flight training.

FINDINGS

A group of jet pilots completing Air Combat Maneuvering (ACM) training in the F-14 were tested on this battery. The few significant correlations found between the test measures and ACM performance were of insufficient quantity or strength to establish that such a battery would reliably predict ACM performance. This could have been due to the homogeneous nature of the subject group in terms of pilot skills and abilities.

RECOMMENDATIONS

Research of this type utilizing this test battery should be continued. Differences in test performance among both similar and different pilot-type groups should be investigated. Changes in test structure, equipment, and procedures should be considered. Further research of a longitudinal nature is needed to fully assess the actual predictive ability of these tests in regards to pilot selection and assignment in naval aviation.

Acknowledgments

The authors gratefully acknowledge Mr. Peter Collyer and Ms. Kathy Vogel for their assistance in programming and modifying the cognitive/psychomotor test battery and Mr. Alfred Thomas for his care and diligence in running subjects. The research reported in this paper was completed under the Naval Medical Research and Development Command work unit No. 63706N M0096 M00960.01 0151.

INTRODUCTION

Research is being performed at the Naval Aerospace Medical Research Laboratory (NAMRL) to develop measures of cognitive and psychomotor ability that reliably relate to the simulated and actual flight performance of fleet aviators. The goal is to develop a test battery capable of predicting the operational performance of fleet aviators before posttraining aircraft assignment. Such efforts would aid in the identification of selection criteria for specific fleet aviator communities and support flight training platform assignment (pipeline) decisions.

A number of naval research efforts have been somewhat successful in predicting certain measures of operational aviator performance. Peer ratings obtained during preflight training were useful in identifying both successful and unsuccessful naval aviators during combat in Vietnam (1). During the midsixties (2), a prediction equation based on the evaluation of F-4 Replacement Air Group (RAG) training showed the possibility of reducing RAG attrition by 38%. A combination of psychological tests and actual flight performance measures have been used (3) to successfully predict F-4 carrier landing performance. Also, a regression equation based on the performance of an East coast F-4 RAG reliably predicted performance of a West coast F-4 RAG (4), and an overall experience measure combined with seven undergraduate training grades reliably predicted the overall RAG grade (5). More recently, a set of automated dichotic listening and psychomotor (cursor tracking) test results correlated significantly with some elements of the Air Combat Maneuvering (ACM) performance of a group of Marine F-4 pilots (6).

These studies suggest the possibility of successfully predicting at least some elements of operational performance in various fleet aviation environments. Our approach is to use an automated battery of cognitive and psychomotor tests to predict aviator performance in various operational settings. This report documents the attempt to find significant relationships between performance on this battery and ACM performance on an instrumented training range for a group of F-14 pilots training at NAS Oceana, Virginia.

METHODS

SUBJECTS

Subjects were 66 Navy F-14 pilots who participated in the Fleet Fighter ACM Readiness Program against the VF-43 adversary squadron at NAS Oceana. The age of these subjects was between 24 and 41 years (M = 29.09, SD = 4.11) while the total number of flight hours up to that point was from 350 to 4500 (M = 1472.57, SD = 1068.43).

APPARATUS AND PROCEDURES

Table 1 lists the various tests given, the sequence of their occurence, and the time required to administer each. The entire series was automated using an Apple IIe microcomputer, an Amdek Color I Plus monitor (CRT), and an Apple IIe numeric keypad. All test instructions were presented on the CRT to each subject before the start of each test. The test types are described in the following sections; further details on each type may be found elsewhere (7).

TABLE 1. Sequence, Description, and Operating Times of Automated Tests.

Presentation order	Description	Test times (min) individual/cumulative
1.	Single psychomotor task (PMT), stick only (S)	07 / 07
2.	Single dichotic listening task (DLT)	16 / 23
3.	First multitask (1,2 combined)	05 / 28
4.	Single (PMT), stick & rudder (S&R)	10 / 38
5.	Second multitask (4,2 combined)	05 / 43
6.	Third multitask (4,2 combined)	05 / 48
7.	Single PMT; stick, rudder, & throttle (S&R&T)	07 / 56
8.	Second single PMT (like 7, S&R&T)	04 / 60
9.	Fourth multitask (8,2 combined)	06 / 66
10.	One dimensional compensatory tracking (ODCT)	10 / 76
11.	Absolute difference computation (ADC)	10 / 86
12.	Fifth multitask, ODCT & ADC (10,11 combined)	10 / 96

PSYCHOMOTOR TASK (PMT)

The psychomotor tracking task required subjects to maintain first one, then two, and finally three, randomly displaced cursors on fixed targets on the CRT by manipulating joysticks and foot pedals. Subjects manipulated one Measurement Systems, Inc., joystick (stick or S), located at the front seat edge, with their right hand to control a cursor that moved within the upper two-thirds of the screen just right of center in a backwards (reversed) manner. Locally produced rudder pedals (rudder or R), patterned after those of a Systems Research Laboratories, Inc., psychomotor test device, were used to control a cursor that moved horizontally across the bottom of the screen. Pushing the left pedal moved this cursor to the right while pushing the right pedal moved it to the left. Another Measurement Systems joystick (throttle or T), located on the left seat edge, was manipulated by the subject's left hand to move a cursor vertically on the left side of the screen. The subject pulled this throttle back to move this cursor down and vice versa.

Psychomotor task tests 1, 4, and 7 (see Table 1) were each preceded by a 3-min practice period. Test 4 was divided into two 3-min testing sessions separated by a 20-s rest interval. Psychomotor task test scores were the accumulated total of absolute errors from an ideal target position. For each time-sampling of cursor position, absolute pixel errors were assessed separately along each dimension. The final error score was the sum of all the samplings made across all the dimensions represented in that particular task. This error score was for the total time of that test. This error score total was then divided by the number of minutes of each test analyzed to generate a standard rate of pixel error per minute of test time. The scores of tests 5 and 6 and tests 7 and 8 were averaged for each subject. All PMT error scores from these tests were then transformed by logarithms to base 10 to reduce skewness and to compensate for extreme outliers, thus reducing the complexity of data analyses while retaining all the data points available.

DICHOTIC LISTENING TASK (DLT)

The DLT consisted of a series of letter/digit string sets presented to subjects aurally over binaural headphones via two Jameco JE 520-AP voice synthesizers. Subjects were told which ear to attend to for each trial. Part I was a series of 16 pairs of letters and/or numbers; Part II was a series of 6 more pairs. Subjects were to indicate the digits (0-9) presented to the designated ear in the order of their occurence. Subjects responded with their left hand using a separate keypad placed immediately in front and slightly left of center. The test was preceded by six aural practice trials, which provided immediate performance feedback by visually indicating the letters and digits presented and the subjects' keypad responses. Subjects also completed three multiple-choice questions before beginning the actual test to ensure that they understood the concept of the DLT.

The DLT performance measure was the number of incorrect responses during 12 trials in which a total of 108 correct responses were possible. The number of correct responses made was subtracted from the total possible correct for that particular test, and after adding one, this new adjusted error score was then transformed by using logarithms to base 10 to adjust for both skewness and extreme outliers as was done for the PMT.

MULTITASK PMT/DLT

In all of the multitask conditions, subjects performed both the DLT and PMT simultaneously (a 12-trial DLT and a 4.5-min PMT). During the first multitask condition, subjects performed the DLT and the stick-only PMT(S). During the next two multitask conditions, subjects performed the DLT and the stick-and-rudder PMT (S&R) using their right hand and feet to control the central joystick and the rudder pedals, and their left hand to make keypad responses to the DLT input. During the final multitask condition, subjects performed the DLT and the stick-rudder-and-throttle PMT (S&R&T). In this most elaborate combination, subjects used their right hand and both feet to control the central joystick and the rudder pedals as before but, in addition, used their left hand to control the throttle joystick and voiced their DLT responses using a microphone attached to the headphones. These vocal responses were tape-recorded for subsequent analysis and hand scoring. Before the start of the various multitask combinations, subjects were instructed to perform each task equally well. Performance measures for the PMT and DLT in these multitask conditions were identical to those of the single tasks with PMT errors being recorded for the final 4 min of that test.

ONE-DIMENSIONAL COMPENSATORY TRACKING (ODCT)

For the ODCT, subjects were to center a square cursor within an elongated rectangle. Subjects used their right hand to move a joystick, which was centered on the front seat edge, left and right. The cursor was driven by a forcing function that increased centering effort with distance from center. During this phase of the task, subjects received three 2-min trials separated by 30-s rest periods. The test measure for the ODCT was total pixel deviation error averaged over the three single-task trials.

ABSOLUTE DIFFERENCE COMPUTATION (ADC)

Randomly selected digits between 1 and 9 were presented inside a small square in the middle of the CRT to subjects. Subjects determined the absolute difference between the digit currently displayed on the CRT and the digit previously displayed. The subjects then pressed the corresponding digit-key on the keypad with their left hand as quickly as possible, resulting in the display of another number for compution. Identical digits were not allowed to repeat. Only the digit responses 1, 2, 3, and 4 were possible. Subjects received three 2-min trials separated by 20-s rest periods. Performance measures for the ADC were the number of correct responses made and the average reaction time of these correct responses, both averaged over the three ADC trials.

DUAL-TASK ODCT/ADC

During this phase of testing, subjects performed both the ODCT and the ADC concurrently. The digits for the difference task were centered just above the tracking task. The subjects controlled the tracking task joystick with their right hand and made keypad responses to the difference task with their left hand. Subjects were instructed to perform each task equally well. Subjects received three 2-min trials with each trial separated by 30-s of rest. Test measures for the dual-task ODCT/ADC were the same as those for the single tasks.

OPERATIONAL PERFORMANCE CRITERIA

Aviators who completed the ACM readiness program were scored on objective and subjective performance. Objective performance measures included number of kills, number of losses, kill ratio, time-to-first kill, kill efficiency percentage (both offensive and defensive), overall weapons performance (both number of kills and missiles fired), weapons performance ratio percentage, visual identification (VID) performance (number of kills), and VID performance ratio percentage. Subjective performance measures involved peer evaluations. Such measures were use of environment, techniques, communications, start, game plan usage, lookout, mutual support, aggressiveness, offensive maneuvers, weapons system employment, defensive maneuvers, UHF communications, energy management, mental plot, situational awareness, bugout technique, and reconstruction, as well as an overall score (the average of the previously mentioned scores).

RESULTS

AVIATOR TEST PERFORMANCE

Table 2 presents descriptive statistics of the test performance of the 66 F-14 pilots on these psychomotor and cognitive tests. Due to technical difficulties, the results of the last DLT test (test 9) were not available for analysis and thus were not included in this table. For this subject group, the mean number of errors made on the PMT, regardless of motor complexity level, decreased when the DLT was added. Two-tailed *t* tests for dependent samples showed this difference to be significant for all conditions (all t values > 9.21, all p values < .01) and would indicate that the subjects performed better on the PMT when it was combined with the DLT. In fact, as the DLT was brought on line with the PMT, our microcomputer could not maintain the level of cursor positioning difficulty attained previously due, to processor overload. This overloading also produced a possible reduction in error sampling rate as test complexity increased. An apparent decrease in testing efficiency has been observed before (7) and does not invalidate the usefulness of these results or methodology. Using Friedman two-way ANOVAs (8), we found that the subjects made significantly more errors as PMT complexity increased during both the unitask and multitask conditions (all ANOVA chi-square values > 118.57, all df values = 2, all p values < .001). Because of the confidentiality of the ACM performance data, descriptive statistics on these data have been excluded from this report.

TABLE 2. Descriptive Statistics of Tests.

Test Measure	Mean	SD	n
Unitask DLT	0.72	0.34	66
Multitask DLT w/(S)	0.84	0.34	65
Multitask DLT w/(S&R)	0.81	0.24	65
Unitask PMT (S)	3.03	0.20	66
Multitask PMT (S) w/DLT	2.79	0.15	66
Unitask PMT (S&R)	3.43	0.13	66
Multitask PMT (S&R) w/DLT	3.16	0.14	65
Unitask PMT (S&R&T)	3.59	0.12	66
Multitask PMT (S&R&T) w/DLT	3.43	0.19	66
Single tracking (ODCT)	19.31	7.76	64
Sgle abs diff. (ADC)	58.63	15.63	61
Sgle abs diff. (ADC) RT	2.25	0.48	61
Dual tracking (ODCT)	29.28	11.85	64
Dual abs diff. (ADC)	62.68	15.45	64
Dual abs diff. (ADC) RT	2.12	0.38	64

TEST CRITERION ANALYSIS

Individual Pearson product-moment correlations were performed between the various test battery measures and the ACM performance measures. Of these 435 correlations, only 15 (3.4%) were significant at or above the .05 alpha level. Such a small percentage would indicate that the significant results found were most likely due to chance. Also, these significant correlations did not follow into any logical or explainable pattern, given the almost random arrangement of their positions in the matrix. As a test of overall significance, canonical correlation analysis (9) was performed utilizing all test and ACM performance measures except the redundant overall subjective grade. Canonical correlation was utilized to determine if any linear combination of the predictor variables (test battery measures) would correlate significantly with any linear combination of the criterion variables (ACM measures). The results of this analysis were not significant (canonical R = .90, chi-square = 400.99, df = 420, p = .740). Given the nonsignificant nature of the canonical correlation and the pattern of intercorrelations, we believe that the test battery measures and ACM performance are not statistically related.

For the six training squadrons that donated to the subject pool, no significant differences were found among the squadrons on any of the test battery measures or objective ACM measures utilizing one-way analysis of variance. Any differences found between squadrons on the subjective ACM measures were difficult to interpret due to possible individual squadron differences in grading such measures and thus were not pursued for further analysis. We also found no significant correlations between either age or number of flight hours and the test battery measures.

CONCLUSIONS AND RECOMMENDATIONS

The results of this study indicate virtually no significant relationships between performance on this test battery and ACM performance for this particular type of aviator. Cognitive and psychomotor abilities measured by this test battery did not appear to interact with ACM performance in any significant manner. Whatever test performance variance was found was due mostly to factors different from those producing the variance seen in the ACM scores. Very similar results were found for a group of F/A-18 pilots tested on this battery who were completing Fleet Replacement Squadron (FRS) training (7). Quite possibly, results from such a battery would not correlate significantly with such operational performance measures for any group of experienced pilots. This would most likely be due to the fact that the skill and ability levels found within such a pilot group would have already been greatly equalized across members due to common selection, training, and flight experiences.

From our results with jet pilots, this particular test battery probably would not be useful to predict flight performance at late stages of training (ACM or FRS), although it might be useful during early stages. Further research is needed in which subjects are tested before flight training and then followed throughout their aviation career. Differences in test performance among both similar and different pilot-type groups should continue to be investigated as thoroughly as possible. Results might indicate some significant and reliable differences between the various pilot types, which would be due primarily to differences in innate ability, perhaps unique and necessary to a particular aircraft. Also, changes in test structure to increase testing efficiency and changes in equipment to increase (or at least stabilize) subject effort should be pursued. Such research would aid in the assessment of these tests as it relates to their ability to predict appropriate selection criteria concerning platform assignment decisions.

REFERENCES

1. Rickus, G.M. and Berkshire, J.R., Development of an Aviation Combat Criterion: Preliminary Report, NAMI-1047, Naval Aerospace Medical Institute, Pensacola, FL, August 1968.

2. Bale, R.M., Rickus, G.M., and Ambler, R.K., Replacement Air Group Performance as a Criterion for Naval Aviation Training, NAMRL-1126, Naval Aerospace Medical Research Laboratory, Pensacola, FL, December 1970.

3. Brictson, C.A., Burger, W.J., and Gallagher, T.J., "Prediction of Pilot Performance during Initial Carrier Landing Qualifications." Aerospace Medicine, Vol. 43, pp. 483-487, 1972.

4. Shannon, R.H. and Waag, W.L., Toward the Development of a Criterion for Fleet Effectiveness in the F-4 Fighter Community, NAMRL-1173, Naval Aerospace Medical Research Laboratory, Pensacola, FL, December 1972.

5. Shannon, R.H. and Waag, W.L., The Prediction of Pilot Performance in the F-4 Aircraft, NAMRL-1186, Naval Aerospace Medical Research Laboratory, Pensacola, FL, July 1973.

6. Griffin, G.R., Morrison, T.R., Amerson, T.L., and Hamilton, P.V., Predicting Air Combat Maneuvering (ACM) Performance: Fleet Fighter Readiness Program Grades as Performance Criteria, NAMRL-1333, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1987.

7. Griffin, G.R. and Shull, R.N., Predicting F/A-18 Fleet Replacement Squadron Performance Using an Automated Battery of Performance Based Tests, NAMRL-1354, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1990.

8. Seigel, S., Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill Book Co., New York, NY, 1956.

9. Kerlinger, F.N. and Pedhazur, E.J., Multiple Regression in Behavioral Research, Holt, Rinehart, and Winston, Inc., New York, NY, 1973.