THE GOOD STICK INDEX
A PERFORMANCE MEASUREMENT FOR AIR COMBAT TRAINING

SAMUEL B. MOORE AND CHARLES W. MESHIER
Vought Corporation
and
ROBERT E. COWARD
Air Force Human Resources Laboratory
Tactical Research Branch
Luke Air Force Base, AZ

SUMMARY

Measuring the proficiency with which a pilot performs a basic fighter maneuver is a difficult task. Many parameters come into play ... the least of which is the frequently used term ... skill. A simulated environment provides an easily managed atmosphere in which to develop and design proficiency measurement techniques, or performance measurements, that may eventually be applied in an airborne environment. This paper reports on one candidate system that has been studied and validated based on the expert opinion of instructor pilots providing air combat training to Tactical Air Command (TAC) Pilots.

INTRODUCTION

The Good Stick Index (GSI) was developed to measure student pilot proficiencies in simulated one-on-one air combat. The GSI, as originally formulated by the Vought Corporation, Dallas, consists of four objective performance parameters measured during TAC Air Combat Engagement Simulator (ACES) I training. The four parameters comprising the GSI were subjectively chosen and, from data obtained over many classes, empirically related to derive a predictor of the "winner" or "runner-up" in the double elimination, one-on-one free engagement tournament held at the conclusion of each training session. This derived relationship predicts the winner or runner-up of the double elimination, free engagement "turkey-shoot" with greater than random frequency.

A study was conducted to statistically investigate the predictive ability of the empirically derived relationship as a predictor of the turkey-shoot winner. These analyses were performed using data collected from 12 classes of students in an experiment representative of TAC ACES I training. Additional analyses were performed to obtain correlations of student pilot background data and IP subjective predictions of student ranking relative to GSI scores and actual turkey-shoot rankings.

BACKGROUND

The TAC ACES I training program is conducted by TAC using the Vought Corporation fixed base air combat simulator, Figure 1. The program utilizes two F-4 configured cockpits with full instruments and weapon systems indicators necessary for air-to-air combat simulation in a functional mode. The software modeling is for F-4D and F-4E aircraft flight characteristics. In addition, a MIG 21 is modeled to provide training in dissimilar aircraft engagements.

The Vought Air Combat Simulator, Figure 1, consists of two cockpits, each situated within a 16-foot-diameter spherical screen. Overhead projectors provide dynamic earth-sky horizon scenes and an image of the opponent's aircraft. The aircraft target is a high-resolution color image provided by the Opaque Target Optical Projector System (OTOPS), recently developed by Vought. Each pilot wears a g-suit and sits on a g-seat. As a pilot increases the load factor on the aircraft, the g-suit inflates and the g-seat deflates. The visual display dims as a function of g and time and finally blacks out, with the target image the last to go. The g-seat also provides a buffet cue, beginning as a high-frequency nibble, increasing in amplitude and decreasing in frequency as penetration into the buffet area occurs.

On-line firing and hit cues, engine, aircraft, and weapon sounds add to the realism of the simulated air combat, and a separate bullet model includes the time of flight. Weapon realism extends to the heat and radar missiles, too, as a miss will be scored if the aircraft target exceeds the missile turning/tracking capabilities before the time of flight has elapsed. A pilot scoring system called the GSI measures the relative air combat skills of the pilot.

A unique Instructor Pilot (IP) station that is mobile and that can be operated from alongside the cockpit provides the IP a matchless vantage point. The IP station provides complete control of the simulation,

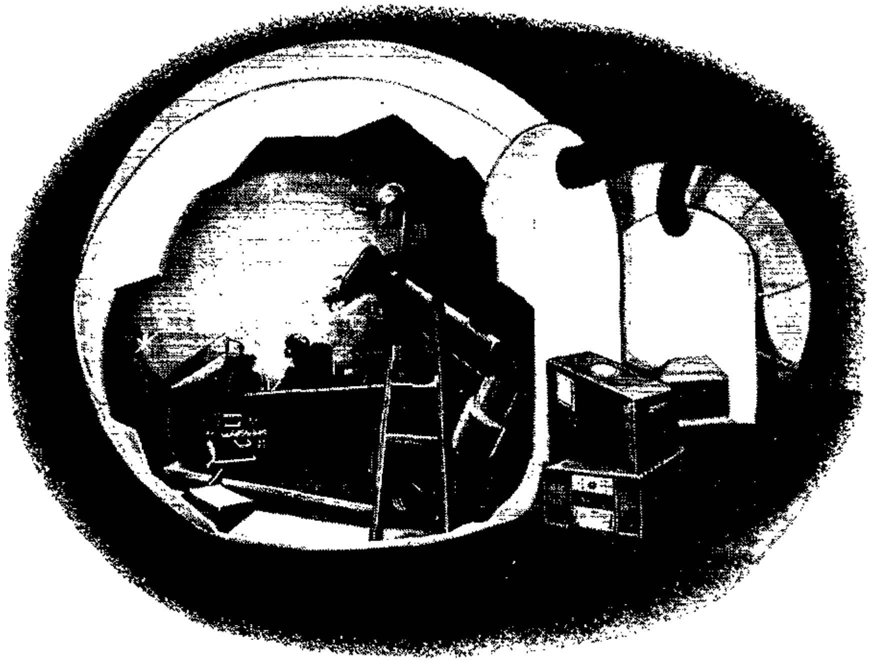


Figure 1. The Vought Air Combat Simulator

including operate, freeze or reset, replay, data recording, video recording, and options to record and play back preprogrammed or canned target trajectories. It also contains the engagement scene which can be recorded on video cassettes, along with the audio from both cockpits and the IP, for subsequent replay and debriefing.

Typically, the TAC ACES I training session is scheduled for one week and consists of eight student pilots and three IPs. Each student accumulates a minimum of 10 hours of classroom and hands-on training in air-to-air combat. Two student pilots train simultaneously in the dual-dome, two-cockpit facility. The student pilot undergoes initial briefings and simulator familiarity sessions on the first day of the 5 days of training. After becoming familiar with the simulator characteristics through the hands-on session, the student is "scored" against a series of canned target maneuvers. The student's initial performance is recorded by computer and stored on magnetic tape.

The final day of training, the fifth day, consists of a second scoring session with each student pilot competing against canned target maneuvers as was initially done on the first day of training. The class training culminates by a double elimination competition, or turkey shoot, where each student competes against the others in one-on-one free engagements until eliminated or a winner is decided.

The data were collected under concisely defined controlled conditions. The study was unique in the sense that the data had to be collected within and from the operational training environment. The collection of data under these conditions also had to be made on a minimum interface and non-interference basis with the ongoing TAC ACES I training program. This requirement precluded the application of experimental controls in a classical sense, as found in a laboratory experiment. As a result, other methods of control were developed to function within the restrictions imposed to provide some assurance as to the fidelity of the data collected and to minimize the effect of undesired variables.

The TAC ACES I students in the study were not aware of the GSI Validation Study and the purposes of data collection. Individual pilot performance data were collected on Monday and Friday of the training week and during the turkey-shoot elimination contest, after completion of the formal training program. In addition, performance data were collected for four of the 12 classes on Wednesday of the training week.

The Chief Instructor Pilot (CIP) predicted each student's performance based on background information (experiences, training, etc.). Also, when formal training was completed on Thursday, the CIPs ranked each student based on his performance. This ranking, prior to the turkey shoot, training system malfunctions, and other history information, were included in the data collection program.

OBJECTIVES

The scope of the investigation was limited to the optimization and validation of the GSI system. The primary products were assessment of the capabilities and limitations of the GSI scores and the determination of the utility of GSI scores as predictors of pilot performance in the simulated free-engagement, turkey-shoot competition.

The original GSI was statistically validated as to its predictive capability by the use of statistical analysis techniques. An improved GSI predictor, using the four subjectively selected parameters, was obtained by discriminant analyses. A further improved GSI predictor was derived from an expanded list of available candidate predictor variables and variable selection techniques. These improved predictors were cross-validated with data acquired from classes outside the experiment. Confidence intervals on the predictors were provided. Standardized discriminant functions were provided to identify the relative contribution of each parameter in the derived predictor equation(s). Student pilot background and subjective data were input with objective data to obtain optimal predictor models.

Subjective rankings of student pilots were compared to the derived GSI predictors and the actual pilot rankings obtained from turkey-shoot results. These interrelationships were described through the use of correlation and variance/co-variance matrixes.

Data from prior classes were used on a random-selected basis to obtain measures of GSI prediction accuracies. These investigations were necessarily limited to the GSI as determined from the four subjectively selected parameters, since other objective data were not on file. A measure of learning effects was obtained by statistically analyzing data from four classes specifically structured to obtain three scoring periods for each student pilot. Measures of individual and group learning were statistically derived as a function of time in training. These learning rates were compared to student pilot performance data.

The reliability of the GSI was determined by calculating confidence intervals of turkey-shoot rank predictions and corresponding confidence levels of the degree of certainty of the predicted value.

ANALYSES

The GSI score was computed from data acquired during the TAC ACES I training of each class, normally on Monday and Friday. During the GSI Validation Study, a third set of GSI data was collected on Wednesday for four of the 12 classes involved. Data are recorded nominally against five canned targets; generally, two of the five are cinetrack and the remaining three are head-on.

The equation defining GSI is,

$$GSI = 4.6\,(70\text{-}MILERR) + 0.86(PANG) + (O/D\text{-}35) + 0.5\,(180\text{-}TTFK) \quad (1)$$

where:

MIL ERR - average mil error over two cinetrack runs while $R < 3{,}000$ ft.

PANG - average percentage of engagement time in pointing angle advantage, $R < 3000$ ft., over two cinetrack runs.

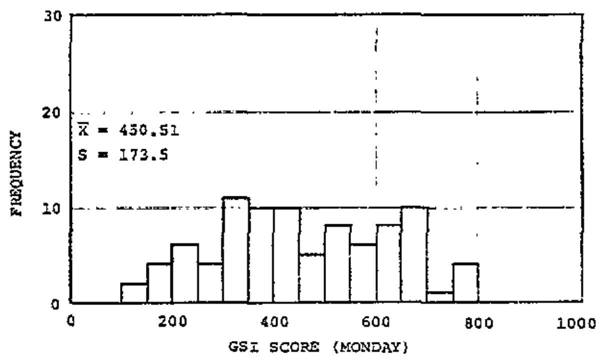
O/D - average ratio of offensive to defensive time against the head-on targets. Offensive time is the time the target aircraft is in the front hemisphere of the piloted aircraft.

TTFK - average time to first kill (seconds) from beginning of run until student achieves first kill against head-on targets with gun or heat missile.

The GSI score has a possible range between zero and 1,000. Each of the four component scores was originally intended to contribute equally to the index. The equation for GSI contains the scaling factors used over the data collection period of this study. MIL ERR, PANG, O/D, and TTFK are referred to as the GSI component scores, or component variables.

The statistical analysis of the Monday and Friday GSI scores and the four GSI component scores collected over the experimental period is presented in this section.

Histograms of the GSI scores and the four GSI component variables (part-scores) were constructed. In general, the score distributions show improvement from Monday to Friday (increase or decrease as appropriate) and the sample standard deviations become smaller (Figure 2).



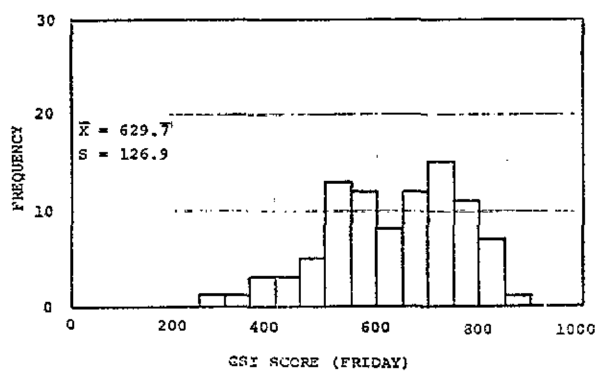


Figure 2. GSI Score Distributions

Correlation coefficients of the four GSI component variables to turkey-shoot rank and fractional wins for both Monday and Friday data are shown in Table 1. The presentation is constructed so that the correlation coefficients for Monday data are shown above the main diagonal of each matrix and for Friday data are below the main diagonal. Relatively strong correlations exist among the component variables indicating non-zero co-variances and thus lack of independence; i.e., possible significant multicollinearities. Correlations between the component variables and turkey-shoot rank and fractional wins are also very weak.

Five three-way analyses of variance were conducted for GSI and the four component variables. The three sources of variation investigated were:

(a) variation between days (Monday and Friday),

(b) variation between turkey-shoot ranks, and

(c) variation between the classes which contained eight students.

Very significant differences existed between Monday and Friday GSI scores, implying, of course, that if GSI measures group learning, a significant increase occurs over the 5-day class period. The other significant source of variation between classes could tend to mask differences between ranks. Scatter diagrams of GSI scores versus turkey-shoot rank showed little significant differences between GSI scores and rank. Analysis of variance for the GSI component variable, average mil error, showed significant differences between ranks but no difference was evident between days. A difference was detectable between classes at the 5-percent confidence level.

The component variable percent PANG showed significant differences between days and between classes. There was no evidence of significance for variation between ranks. The component variable, offensive time, showed significant differences between days. No differences appeared to exist between ranks or classes. For the component variable, TTFK, differences were detected at the 5-percent level between days and between ranks. Differences were not evident between classes.

Table 2 presents a comparison of the best predictor using the Friday only GSI score with random selection and with CIP predictions (CIPPs) made just prior to the turkey-shoot competition. Comparisons were made at four levels of detail as to the outcome of the turkey-shoot. The four levels are defined as follows:

1. Four Groups - Proper placement into the proper turkey-shoot quartile, i.e., 1 or 2 in the first group; 3-4 in the second group, 5-6 in the third group and 7-8 in the fourth group.

2. Upper Half of Class - Proper placement of students in the top four turkey-shoot ranks in those ranks; i.e., 1, 2, 3, 3.5 or 4.

3. Winner and Runner-Up - Proper placement of the winner or runner-up in the winner/runner-up group.

4. Winner - Proper identification of the actual turkey-shoot winner.

TABLE 1 - GSI CORRELATION COEFFICIENTS

(MONDAY above the diagonal; FRIDAY below the diagonal)

	T.S. RANK	AVG.MIL ERR	% PANG	% OFF TIME	TTFK
T.S. RANK	1	.1254	-.1318	-.0270	.1512
AVG.MIL.ERR.	.0200	1	-.0891	-.1915	.1650
% PANG	.0313	-.3071	1	.2107	-.2868
% OFF TIME	-.2761	-.0951	.0007	1	-.5430
TTFK	.2817	.0559	-.1557	-.6052	1

(MONDAY above the diagonal; FRIDAY below the diagonal)

	FRACT. WIN	AVG.MIL ERR	% PANG	% OFF TIME	TTFK
FRACT. WIN	1	-.1355	.1759	.0261	-.1218
AVG.MIL.ERR.	-.0083	1	-.0891	-.1915	.1650
% PANG	.0289	-.3071	1	.2107	-.2868
% OFF TIME	.2866	-.0951	.0007	1	-.5430
TTFK	-.2748	.0559	-.1557	-.6052	1

TABLE 2 -

A COMPARISON OF FRIDAY GSI RANK PREDICTIONS WITH CHIEF INSTRUCTOR PILOT (CIPP) AND RANDOM SELECTION

GROUPINGS		RANDOM SELECT.	CIPP	GSI RANKING (FRI. SCORE)
FOUR GROUPS (1-2, 3-4, 5-6, 7-8)	• NO. CORRECT PREDICT.	-	21	26
	• TOTAL NO. PREDICT.	-	67	90
	• % CORRECT PREDICT.	25%	31.3%	28.9%
	• 95% CONFIDENCE INT.	-	20.2-42.5	19.5-38.3
UPPER HALF OF CLASS (1,2,3-4)	• NO. CORRECT PREDICT.	-	24	27
	• TOTAL NO. PREDICT.	-	34	46
	• % CORRECT PREDICT.	50%	70.6%	58.7%
	• 95% CONFIDENCE INT.	-	55.2-85.9	44.5-72.9
WINNER & RUNNER-UP (1, 2)	• NO. CORRECT PREDICT.	-	6	9
	• TOTAL NO. PREDICT.	-	17	23
	• % CORRECT PREDICT.	25%	35.3%	39.1%
	• 95% CONFIDENCE INT.	-	12.6-58.0	19.2-59.1
WINNER (1)	• NO. CORRECT PREDICT.	-	1	3
	• TOTAL NO. PREDICT.	-	9	12
	• % CORRECT PREDICT.	12.5%	11.1%	25.0%
	• 95% CONFIDENCE INT.	-	0-31.6	0-49.5

The CIPPs were made for 67 out of a possible 90 CIPPs. The random-selection probabilities were determined under the assumption of independent random assignment of students to the turkey-shoot position.

Four entries are provided for CIPP and GSI ranking predictors for each of the four groupings. These provide basic data on the actual predictions. For example, for CIPP and the "four-groups" grouping, the CIPs properly placed 21 out of 67 predictions in the correct groupings (1-2, 3-4, 5-6, or 7-8); thus, 21 of 67 or 31.3 percent were correctly classified. Ninety-five percent confidence limits were calculated using these data and were determined to be 20.2 percent and 42.5 percent (1).

Each CIPP and GSI ranking prediction was subjected to a test of the hypothesis that it is equal to or better than random selection (2). The CIPP for the upper half of the turkey shoot was found to be significantly better than random selection for winner and runner-up also at the 5-percent confidence level. All other predictions were found not to be significantly different from random prediction at the 5-percent level. Table 3 provides the levels of significance at which differences would be assumed to exist.

TABLE 3 - APPROXIMATE RISK LEVEL AT WHICH DIFFERENCES CAN BE ASSUMED TO EXIST

GROUPINGS	CIPP	GSI RANKING
FOUR GROUPS	15%	18%
UPPER HALF	5%	13%
WINNER & RUNNER-UP	26%	5%
WINNER	36%	20%

It can be concluded; CIPPs can classify students as to whether or not they will place in the upper half of the turkey shoot with up to 86-percent accuracy. A simple GSI ranking scheme can correctly predict turkey-shoot winner and runner-up classification about 39 percent of the time. For other predictions investigated, the two predictors appear to be no better than random selection.

The Monday and Friday GSI scores, GSI component variables, an expanded set of candidate predictor variables, and demographic data were subjected to a series of discriminant analyses using the sub-program DISCRIMINANT available as part of the SPSS package (3). The capabilities of this program were useful in the development of improved predictor equations from the available data. The purpose of the analysis was to statistically derive optimal models which predict turkey-shoot rank from data collected during the 12-class TAC ACES I experiment. The derived models used the Wilks' Lambda variable selection criteria to select the best candidate predictor variables from those available. The models derived are optimal within the constraints of the analysis but are not necessarily maximal. A maximal predictor model could only be achieved if all possible models were considered.

Discriminant analysis considers a desire to statistically distinguish between two or more defined groups using information available from sample data. The groupings of interest were defined from turkey-shoot rank. In a normal class of eight student pilots, there are at least five distinguishable turkey-shoot groupings. These are, in order from most favorable to least favorable outcome: winner (1), runner-up (1), third eliminators (2), second eliminators (2), and first eliminators (2).

The primary objective of the analysis was to develop predictor algorithms for turkey-shoot winners; therefore, the groupings considered were structured to investigate the level of detail at which winners could be predicted from available data. Winners were classified in two ways. One winner class was the absolute winner, or undefeated student, in the turkey shoot. A second winner class was the winner and runner-up. This grouping scheme was used with some limited success in earlier Vought investigations which employed only the Friday GSI score as the predictor variable. In all, four different grouping schemes (winners, winners and runners-up together, the upper half of the class, and quartile groupings) were defined and investigated.

The analysis to obtain an improved GSI was conducted in four parts, each part being defined by the candidate predictor variable set to be used. The first analysis used Monday and Friday GSI scores as candidate predictor variables. This analysis provided a measure of the best prediction capability of the GSI score.

The discriminant analysis correctly classified most of the true winners but incorrectly classified a relatively large number of non-winners as winners. By using indicators more complex than the composite GSI score, it was possible not only to correctly classify winners a fairly large percent of the time, but also to greatly reduce the incidence of non-winners improperly placed into the winners' group.

In the second analysis, the four component variables (or part scores) from which GSI is calculated were used instead of the composite GSI scores. The DISCRIM program was then allowed to select from these eight component variables (four for Monday and four for Friday) the best predictor variables for each of the four classification schemes. The eight variables are defined in Table 4 which shows that DISCRIM was selective and did not use all available data to define the optimal prediction (classification) equations.

TABLE 4 - MONDAY AND FRIDAY GSI COMPONENT VARIABLES AND VARIABLE SELECTION BY DISCRIMINANT GROUP

VAR. DESIG.	GROUP I - Winners; GROUP II - Others	GROUP I - Winners & Runners-Up; GROUP II - Others	GROUP I - Winners, R.U., & 3rd Elim.; GROUP II - Others	GP. I - Win. & R.U.; GP. II - 3rd Elim., GP. III - 2nd Elim.; GP. IV - 1st Elim.	VARIABLE DEFINITION
X1		X		X	AVERAGE MIL ERROR FOR FRIDAY
X2		X			PERCENT TIME IN PANG FOR FRIDAY
X3	X		X	X	PERCENT OFFENSIVE TIME FOR FRIDAY
X4		X	X		TIME TO FIRST KILL ON FRIDAY (SECONDS)
X5		X		X	AVERAGE MIL ERROR FOR MONDAY
X6					PERCENT TIME IN PANG FOR MONDAY
X7					PERCENT OFFENSIVE TIME FOR MONDAY
X8	X		X	X	TIME TO FIRST KILL ON MONDAY (SECONDS)

Results of the Discriminant Analysis

In the discriminant analysis where Monday and Friday GSI scores are the predictor variables, members of the first group are correctly classified in the order of 60 percent of the time, but many non-first group students are classified incorrectly in the first group. The lack of discriminant power was evidenced by low values of the canonical correlation coefficients of the respective discriminant functions; i.e., between 0.120 and 0.218.

The eight GSI component variables (four for Monday GSI component scores and four for Friday GSI component scores), when used as candidate predictor variables, resulted in better predictive capabilities than in the GSI score analysis.

Candidate predictor variables were developed from an objective data set collected during the Monday and Friday GSI scoring session but not previously analyzed.

In the expanded data-set analysis, over 80 predictor variables were available for consideration as candidates for the analysis. An initial screening of the complete list was necessary to reduce the number of variables to an acceptable size. This screening was accomplished by correlating all variables with turkey-shoot rank and then selecting the 40 variables from the list with the greatest correlation coefficients. Table 5 shows those objective variables which were selected by DISCRIM as the best turkey-shoot rank predictors. In this table, the predictor variables were separated by day of data collection. The discriminant classification schemes by which each were used is also indicated. Use of this expanded list of candidate variables appears to have generally improved the winner prediction capability.

The canonical correlations of the discriminant functions of the analyses greatly increased over analogous functions in the previous analysis, indicating increased capability to discriminate between groups. This increased discriminant capability is at the cost of increased complexity in the number of variables required and the complexity of calculations. The classification functions provided optimal predictors for the objective data analyses and included the best predictor variables consistent with the Wilks' Lambda variable selection criteria. The analyses provided correct classification into Group I on the order of 80 percent; however, a fairly large number of non-Group I members were misplaced.

The analysis used as candidate predictor variables all of the objective data set plus seven demographic variables. Comparison of the results with the expanded objective data analysis showed predictions to be as good or better than before. Mis-classification into Group I was reduced in three of the four classifications, and correct classification into Group I was improved slightly in two of the four classifications. Evidence of this improved discrimination was provided by improvements (increases) in the canonical correlations of the discriminant functions.

In the first discriminant classification scheme (Group I - Turkey-Shoot Winners, Group II - Other), the number of predictor variables required to maintain a constant correct classification rate was reduced from 11 to 10 by inclusion of demographic data.

Inclusion of the demographic data with the expanded objective set caused several Monday objective variables to be excluded in

TABLE 5 - SELECTED OBJECTIVE DISCRIMINANT VARIABLES

VAR. DESIG.	WINNER VS OTHERS	WIN/R.U. VS OTHERS	UPPER 1/2 VS LOWER 1/2	FOUR GROUPS	VARIABLE DEFINITIONS
M4			X		TOTAL FUEL USED (LBS./AVG./HEAD-ON)
M8	X				TOTAL ROUNDS FIRED (NO. TOTAL/HEAD-ON)
M9				X	TOTAL TIME SR LT 1500 (SEC-AVG./CINETRACK)
M10		X			AVG. MIL. ERROR SR LT 3000 (MILS-AVG./CINETRACK)
M11				X	TIME TO PANG (SEC-AVG./CINETRACK)
M12	X				TIME TO FIRST KILL (SEC-AVG./HEAD-ON)
M14		X			DELTA ENERGY STATE - CINETRACK (INT. - END/CTK)
M16	X				TOTAL NO. HITS - CINETRACK (HITS/CTK)
M20			X		TOTAL TIME IN R-MIS ENVELOPE - CTK (TIME/CTK)
M22				X	TIME TO GUN ENVELOPE - CINETRACK (TIME/CTK)
M24		X			TOTAL TIME IN GUN ENVELOPE - CTK (TIME/CTK)
M25			X	X	TOTAL TIME IN GUN ENVELOPE - HEAD-ON (TIME/H.ON)
M29	X	X			HIT/MISS HEAT - MIS. SCORE - H.ON (H*(H+M)/H.ON)
M32	X	X		X	HIT/MISS GUN SCORE (H*TOTAL RDS/H.ON)
F1			X	X	MAX G'S (MAX/SERIES)
F11	X				TIME TO PANG (SEC-AVG./CINETRACK)
F16				X	TOTAL NO. HITS CTK (HITS/CTK)
F18	X	X	X	X	TOTAL TIME IN H-MIS. ENV. CTK (TIME/CTK)
F22	X				TIME TO GUN ENVELOPE CTK (TIME/CTK)
F23	X			X	TIME TO GUN ENVELOPE H.ON (TIME/H.ON)
F25			X	X	TOTAL TIME IN GUN ENVELOPE H.ON (TIME/H.ON)
F27	X	X		X	G-SPREAD H.ON (MAX. G-MIN G)
F29	X	X	X	X	HIT/MISS H-MIS SCORE H.ON (H*(H+M)/H.ON)
F30			X		HIT/MISS R-MIS SCORE H.ON (H*(H+M)/H.ON)

the resulting algorithm.

Comparison of Prediction Results

Table 6 summarizes the predictive capabilities of the major predictor models presented. The table also includes approximately 95 percent confidence limits on the prediction rates (4). Note that the confidence limits are approximate and use the normal approximation to the binomial. This requires a relatively large sample size. For predictions of the winner (the last row of the table), sample size is nine or 12.

Given the predictor models developed using discriminant analysis, it is necessary to test these models using data collected outside the experimental data set. The purpose of these tests is to determine if the predictability of the developed models is retained using predictor variable data not used in the calculation of the parameters or in the selection of the predictor variables. In the analysis performed, there is evidence that the parameters selected are very sensitive to the particular data set used in their estimation and to the definition of the discriminant groups. The values of the parameter estimates are also probably quite sensitive to the data set used.

A very limited test analysis using data obtained prior to this study has been conducted on the predictor models developed from Monday and Friday GSI scores and Monday and Friday GSI component variables. No additional test analysis was conducted.

PSYCHOMETRIC AND EDUMETRIC DATA ANALYSIS

Edumetrics is defined herein as the measurement of an individual's gains from training experiences by the quantitative assessment and analysis of performance data, to include individual and group data. Edumetrics is shown to be concerned with measures of learning performance in contrast to psychometrics, which is concerned with the measurement of individual differences (i.e., measures of individual innate abilities and traits).

Individual and group performance data were recorded for the 80 subjects in this study. The mean GSI performance scores for the Monday and the Friday data sessions were calculated for each of the 12 classes. Two least-squares linear-trend lines were computed, using the Monday GSI scores and the Friday GSI scores.

Four of the 12 TAC ACES classes in this study were subjected to separate analysis. In addition to the normal TAC ACES Monday and Friday data collection sessions, GSI performance data were recorded on Wednesday of the training week. This yielded three sets of performance data for each of the four classes. Scatter diagrams, linear and quadratic curves, and frequency distributions were constructed. A total of 81 data points

TABLE 6 - COMPARISON OF PREDICTION RESULTS

GROUPINGS		RANDOM SELECTION	CIPP	GSI RANKING (FRI. SCORE)	DISCRIMINANT ANALYSIS: GSI SCORE MON. & FRI.)	DISCRIMINANT ANALYSIS: GSI PRED. VAR.	DISCRIMINANT ANALYSIS: EXP. LIST	DISCRIMINANT ANALYSIS: EXP. LIST +DEM. VAR.
Four Groups (1-2, 3-4, 5-6, 7-8)	. No. Correct Pred.	-	21	26	N/A	36	56	57
	. Tot. No. Pred.	-	67	90		90	89	89
	. % Correct Pred.	25%	31.3%	28.9%		40.0%	62.9%	69.3%
	. 95% Conf. Int.	-	20.2 - 42.5	19.5 - 38.3		29.9 - 50.1	52.9 - 73.0	54.1 - 74.0
Upper Half of Class (1,2,3,4)	. No. Correct Pred.	-	24	27	27	27	36	37
	. Tot. No. Pred.	-	34	46	46	46	46	46%
	. % Correct Pred.	50%	70.6%	58.7%	58.7%	58.7%	78.3%	80.4%
	. 95% Conf. Int.	-	55.2 - 85.9	44.5 - 72.9	44.5 - 72.9	44.5 - 72.9	66.3 - 90.2	69.0 - 91.9
Winner & Runner-Up (1, 2)	. No. Correct Pred.	-	6	9	14	15	19	19
	. Tot. No. Pred.	-	17	23	23	23	23	23
	. % Correct Pred.	25%	35.3%	39.1%	60.9%	65.2%	82.6%	82.6%
	. 95% Conf. Int.	-	12.6 - 58.0	19.2 - 59.1	40.9 - 80.8	45.8 - 84.7	67.1 - 98.1	67.1 - 98.1
Winner (1)	. No. Correct Pred.	-	1	3	8	9	10	10
	. Tot. No. Pred.	-	9	12	12	12	12	12
	. % Correct Pred.	12.5%	11.1%	25.0%	66.7%	75.0%	83.3%	83.3%
	. 95% Conf. Int.	-	0 - 31.6	0 - 49.5	40.0 - 93.3	50.5 - 99.5	62.2 - 100	62.2 - 100

were used to fit linear and quadratic least-square lines for all four classes in the sample. Both the linear and the quadratic equations developed approximate the centroid of the mass of data points for each pilot, Figure 3. Linear and quadratic lines fit the data well. Objective measures of these fits are shown in the edumetric analysis. The quadratic curve is preferred in describing the data because it approximates true learning rates, which tend to be non-linear as a function of time. Here,it specifically shows a higher rate of learning during the early phases of training and a lower, slower rate during the final training phases. The distribution of the GSI scores by day of training are shown characterized by normal distributions in Figure 4. It can be seen that the mean ($\overline{X}$) GSI scores improved with length of training.

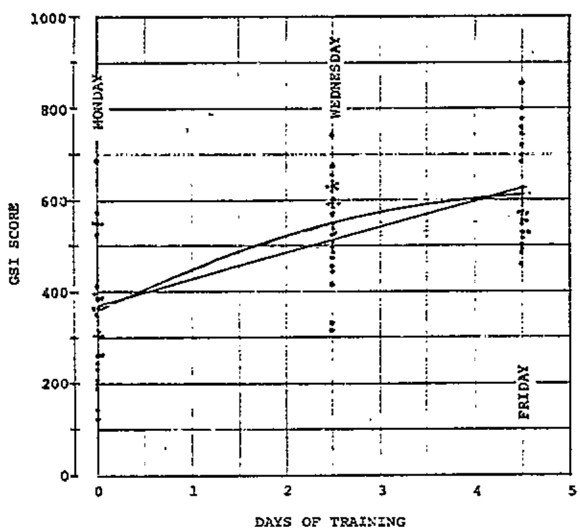


Figure 3. Scatter Plot of GSI Scores.

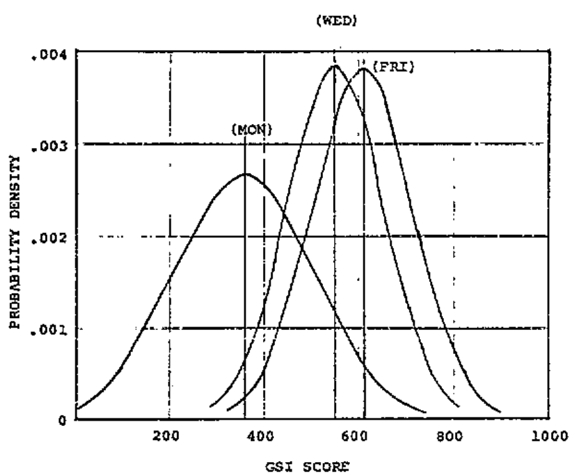


Figure 4. GSI Score Densities

The standard deviation of the scores decreased as length of training increased. This would indicate the effects of learning. The reduced variability in the Wednesday and Friday Standard Deviation values suggests that the subjects were using their experiences gained during the first 2-1/2 days of training and calibrating their performance responses to the expected and anticipated performance of the canned targets.

The degree of individual change in performance score for each subject in this sample over the 4.5-day training week indicates the individual subjects had a mean performance score (GSI) improvement of 61.3 percent for the 27 subjects in the sample.

Using performance data collected on Wednesday for four of the 12 classes, the method of analysis was to fit a straight line and a quadratic curve through the data. The objective was to ascertain the general trend n GSI scores as a measure of group learning rates as the classes progressed.

A scatter diagram of the GSI scores versus days of training shows the linear and quadratic least-squares curves fit through the data. Both curves can be seen to fit well through the central regions of the data for each day. Each fit shows the general trend of GSI Score increasing with days of training.

A further point of interest is the actual normality of the distributions of the GSI scores being analyzed by day; that is, is there any reason to doubt that a given set of scores is normally distributed? The Kolmogorov-Smirnov (K-S) test of goodness of fit was applied to GSI scores for each day. (5). The scores were found to be normally distributed at the percent significance level for each of the three sets of GSI scores.

CONCLUSIONS AND RECOMMENDATIONS

As a candidate technique for simulated ACM performance measurement, the GSI has been analyzed using data from a control group to validate the potential for predicting and ranking students in the TAC ACES I Training Program. The GSI, developed from empirical data, predicts student pilot performance comparable to the predictions of ACM experts ... the instructor pilots. This study examined available historical data demonstrating the value of documenting professional background (experience), student performance results, and instructor evaluations as baseline data for formal training courses.

Failure to collect ample data on training experiences can impact on future potential to account for training expenditures (accountability) and end results (effectiveness).

The GSI has been shown to predict student pilot ranking, in the TAC ACES I training program, with 75 percent accuracy. By adding available objective and subjective data, the prediction improved 80 percent. Such measures, in a simulated environment, have potential training value in the dynamic setting of airborne engagements using the data transmission capabilities of an automated range.

It follows that if a measurement can predict training outcomes, then the measurement parameters must be good indicators of performance. Systems like the GSI should be applied to available simulators and to an air-combat maneuvering range toward developing an objective measure of transfer of ACM training to the airborne mission. This next step is the logical order for moving automated performance measurement of ACM skills to the live-battle scene.

REFERENCES

1. Ostle & Mensing. Statistics in Research, (3rd ed.). Ames: Iowa State University Press, 1975, 100-101.

2. Ostle & Mensing. Statistics in Research, (3rd ed.). Ames: Iowa State University Press, 1975, 129-133.

3. Nie. Statistical Package for the Social Sciences (SPSS), (2nd ed.). New York: McGraw Hill, 1975, 434-462.

4. Ostle & Mensing. Statistics in Research, (3rd ed.). Ames: Iowa State University Press, 1975, 100-101.

5. Ostle & Mensing. Statistics in Research, (3rd ed.). Ames: Iowa State University Press, 1975, 489-490.

ABOUT THE AUTHORS

MR. SAMUEL B. MOORE is Project Manager, Simulation and Training Systems, at Vought Corporation in Dallas. He is currently responsible for Vought research and development work in simulator application to the proposed next generation Navy Training System, VTXTS. He has 23 years with Vought, with background in experimental aerodynamics, structures tests, systems tests, and flight simulation. He is an associate fellow of AIAA and a licensed professional engineer. He holds a B.S. degree in science from Mississippi State University and an M.S. degree in aeronautical engineering from Southern Methodist University.

MR. CHARLES W. MESHIER, a former Naval Aviator and retired Navy Captain, is Project Manager, Air Combat Training, Vought Corporation, Dallas, with 12 years of experience in the Simulation and Training Systems Office. His aviation experience reaches back to the F6F Hellcat, F4U Corsair, and early helicopters. He also commanded Heavy Attack Squadron Six and served in the Bureau of Aeronautics as the Rotary Wing Aerodynamicist. He is an AIAA associate fellow and a registered professional engineer. He holds a B.S.E. from the U.S. Naval Academy post-graduate school and an M.S.E. from Princeton.

MR. ROBERT E. COWARD is the Senior Research Psychologist for the Tactical Research Branch of the Flying Training Division, AF Human Resources Laboratory, Luke AFB, Arizona. Since 1977, Mr. Coward has been performing specialized research oriented to the requirements of flying training in a tactical environment. Before this, Mr. Coward was assigned to Headquarters, Aerospace Defense Command where he served as Chief of Instructional Technology Division. He has a B.A. degree in psychology from the University of Southern Colorado and an M.A. degree in psychology from the University of Colorado.