AD-A199 004

IDA PAPER P-2085

RELATING FLYING-HOUR ACTIVITY TO THE PERFORMANCE OF AIRCREWS

Stanley A. Horowitz
Colin P. Hammon
Paul R. Palmer

December 1987

DTIC
ELECTE
SEP 1 5 1988

E

Prepared for
Assistant Secretary of Defense,
Force Management and Personnel

INSTITUTE FOR DEFENSE ANALYSES
1801 N. Beauregard Street, Alexandria, Virginia 22311

88 9 14 208   IDA Log No. HQ 88-33010

AD-A 79 ...4

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS			
UNCLASSIFIED				
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT			
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	Approved for Public Release; distribution unlimited			
4. PERFORMING ORGANIZATION REPORT NUMBER(S) IDA Paper P-2085	5. MONITORING ORGANIZATION REPORT NUMBER(S)			

6a. NAME OF PERFORMING ORGANIZATION Institute for Defense Analyses	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION OUSDRE (DoD-IDA Management Office)		
6c. ADDRESS (City, State, and Zip Code) 1801 N. Beauregard St. Alexandria, VA 22311		7b. ADDRESS (CITY, STATE, AND ZIP CODE) 1801 N. Beauregard St. Alexandria, VA 22311		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION ASD, Force Mgt. and Personnel	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA 903 84 C 0031		

8c. ADDRESS (City, State, and Zip Code) Room 2B284, The Pentagon Washington, DC 20301	10. SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT	PROJECT NO.	TASK NO. T-L7-516	ACCESSION NO. WORK UNIT

11. TITLE (Include Security Classification)
Relating Flying-Hour Activity to the Performance of ~~Aircraft~~ *Aircrews*

12. PERSONAL AUTHOR(S).
Stanley A. Horowitz, Colin P. Hammon, Paul R. Palmer

13. TYPE OF REPORT Final	13b. TIME COVERED FROM Jun 87 TO Dec 87	14. DATE OF REPORT (Year, Month, Day) ~~February 1988~~ *December 1987*	15. PAGE COUNT 30

16. SUPPLEMENTARY NOTATION

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Aircrew Performance, Flying Hours, Training, Readiness, Performance, Proficiency, Aviation Training

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

    This paper assesses the feasibility of developing quantitative relationships between the capability of aviation units to perform their assigned missions and the level of resources available to them for training. It responds to skepticism about the impact of reductions in the number of flying hours available for aircrew training on aircrew performance.

    The relatively sparse literature that develops statistical links between aircrew performance and flying hours is reviewed. Quantitative evidence of the magnitude of these links is discussed, and a model for relating flying-hour activity to aircrew performance is developed. The model posits that additional flying enhances proficiency in two ways: through the short-run honing of skills and through the long-run development of mastery. Available data on aircrew performance and on training histories are described.

    The paper finds that statistical analysis of historical information can successfully quantify the effects of training and experience on aircrew proficiency and safety. Additional analyses of the data that exist to support such quantification are proceeding.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL Stanley A. Horowitz	22b. TELEPHONE (Include Area Code) (703) 845-2450	22c. OFFICE SYMBOL IDA/CARD

DD FORM 1473, 84 MAR   83 APR edition may be used until exhausted.
 All other editions are obsolete

IDA PAPER P-2085

RELATING FLYING-HOUR ACTIVITY
TO THE PERFORMANCE OF AIRCREWS

Stanley A. Horowitz
Colin P. Hammon
Paul R. Palmer

December 1987

IDA

INSTITUTE FOR DEFENSE ANALYSES

PREFACE

This paper presents the results of the first phase of an investigation into the relationships between the flying hour history of aircrews and measures of their proficiency. It reports the feasibility of developing such relationships.

The purpose of the analysis is to assist personnel in the Office of the Secretary of Defense and in the military services to demonstrate that proposed changes in funding for flying hours, the major vehicle for aircrew training, will cause quantifiable changes in performance.

CONTENTS

FIGURES

TABLES

I. INTRODUCTION AND SUMMARY

The purpose of this study is to develop quantitative relationships between the capability of aviation units to perform their assigned missions and the level of resources available to them, using information on the performance of aircrew personnel. It is meant to directly address concerns voiced by the General Accounting Office (GAO) and skepticism displayed by Congress about the impact of cuts to the flying-hour programs of the services.

The study is designed to be performed in three phases. This paper reports on the results of the first phase. Our goal here is to show that it is feasible to build the kinds of quantitative relationships between capability and resources that we seek to develop. Phase two will be designed to produce illustrative examples of such relationships. If the first two phases are successful, phase three is meant to initiate a broad research effort covering all the services and a wide range of aircraft types.

Our general approach is statistical. We want to use statistical techniques to examine historical data in order to relate indicators of proficiency, including indicators of safety, to training histories. This requires data on the output of the training process -- proficiency -- as well as data on the inputs -- principally flying-hour histories, but including, where possible, information on the use of simulators. It also requires a conceptual framework for linking the two.

The rest of the paper is divided into six sections. The first describes the concerns that motivate this study. The second section reviews the sparse but interesting body of literature relating aspects of aircrew performance to flying hours. After that, a model for relating flying-hour activity to aircrew performance is developed. This is followed by a description of the data on aircrew performance that have been identified in our initial explorations and a discussion of our plans to analyze that data. The paper ends with conclusions.

We find that historical information can be successfully used to quantify the effects of training and experience on aircrew proficiency and safety. We also find that data exist to support such quantification. All the necessary data to perform two case studies relating

1

flying hours to proficiency measures have been obtained. Analyses of these data are proceeding and more data sets are being developed.

II. BACKGROUND

All of the military services spend a considerable amount of money flying aircraft in peacetime. This includes expenditures on aviation fuel, on spare parts and on full-time maintenance personnel. Most of this flying is for the purpose of maintaining and upgrading the proficiency of aircrew personnel. Recently doubts have been raised about the extent to which changes in levels of flying-hour activity would increase or decrease the ability of aircrews to effectively perform the tasks for which they are being trained, and which they might have to execute in a hostile environment.

A recent report on aircrew training by the General Accounting Office (GAO) notes that the Tactical Air Command (TAC) and the Strategic Air Command (SAC) base their criteria for determining how many flying hours are needed to maintain and enhance pilot and crew proficiency largely on the judgment of experienced pilots. It finds that the Air Force does not have a system for aggregating and analyzing data used as the basis for its professional judgments [1]. The report concludes that there is a need to develop and maintain a system for using objective data to assess the benefits pilots and aircrews receive from different levels of flying.

GAO's findings reflect widespread Congressional skepticism about the validity of the requirements for flying hours stated by each of the services. This skepticism has manifested itself in continuing pressure on the flying-hour budget. Congress has not been satisfied with the services' responses to requests that they show the implications of changes in flying hours for aircrew performance. Traditionally these responses rely heavily on the methodology used to develop flying hour programs. Figure 1 presents an overview of that methodology.

Every aircrew for each type of aircraft has a set of missions to execute and a set of tasks that must be performed to execute them. The frequency with which these tasks must be repeated to maintain proficiency is based on informed professional judgment and observation. These tasks and frequencies combine to form the training syllabus. Required training programs are built from the number of hours needed to execute the syllabus and the

3
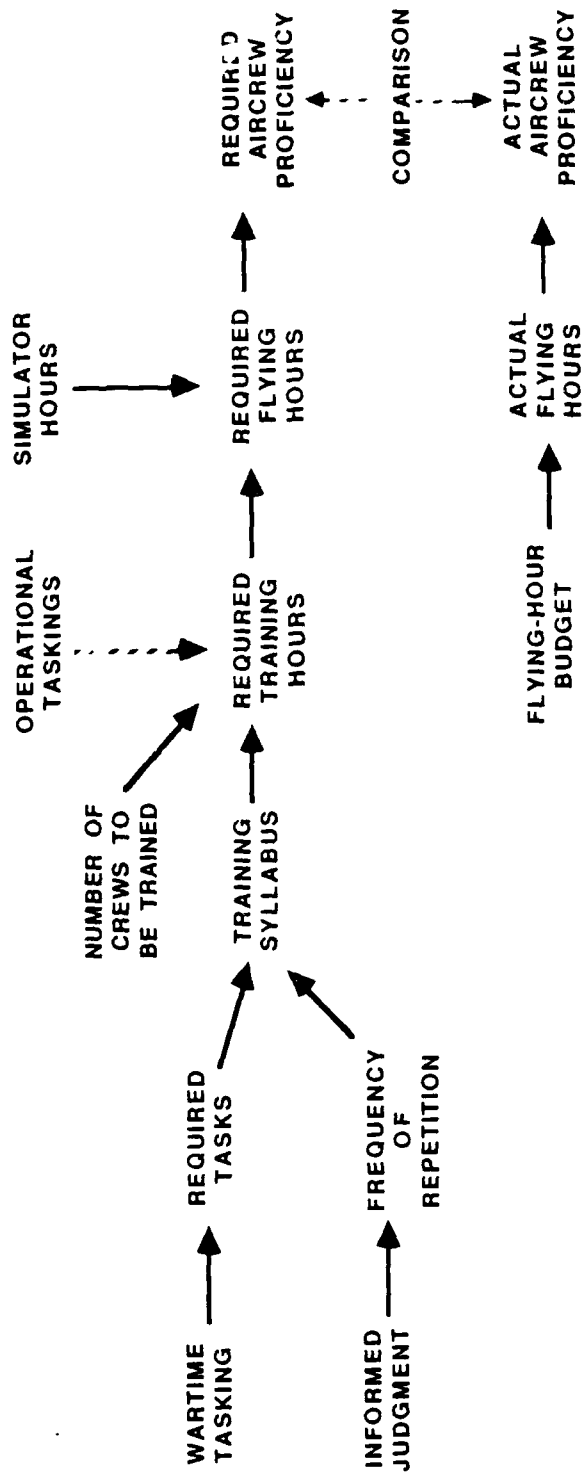
FIGURE 1. AN OVERVIEW OF THE FLYING HOUR ISSUE

number of crews to be trained. The requirement to do non-training-related operational tasks ought to be added in, but often is not. Some training is performed in simulators. That portion of the training program for which simulators are not available or felt to be not suitable determines the flying-hour requirement. Flying this number of hours is expected to yield the needed level of aircrew proficiency. Generally the required number of flying hours determined by this methodology exceeds the actual number that can be bought with the flying-hour budget. Presumably the greater the difference between the actual and required programs, the greater the difference between actual and required aircrew proficiency.

Questions concerning the implications of changing the flying-hour budget can be answered by reference to the training syllabus. Less flying implies that more of the required tasks will not be fully trained for and that aircrews will not be qualified to perform as many of their missions. The weakness of this estimate of the impact of reduced flying is that it is not validated by explicit reference to the actual performance of any group of aviators. Except for a few isolated cases, the services have not been able to point to two groups of aircrews and demonstrate that the group that flew more could perform better.

A reason for this situation is that making such a comparison requires data on indicators of the military performance of aircrews. It is hard to measure military performance. Researchers have noted "the lack of information on job performance resulting from training [2]." A recent paper, however, showed this lack to be less pervasive than is widely believed. All the services go to considerable effort to develop indicators that are closely related to military effectiveness [3]. They are generally used for management purposes in the field. They are usually not forwarded to higher headquarters or used to assess the effectiveness of manpower, personnel and training policies.

III. SUMMARY OF PREVIOUS RESEARCH

Most of the existing literature on the relationship between flying hours and aircrew performance has been developed using Navy data, though one particularly ambitious study examined tactical bombing performance in the Air Force. These studies use a diverse set of variables to reflect performance. The performance indicators include final grades on Operational Readiness Evaluations (OREs), boarding rates for carrier-based aircraft, carrier landing grades, accident rates and bombing accuracy. Some of the analyses focus on recent flying hours, while some examine the total number of flying hours accumulated over the course of a career.

A. RECENT FLYING HOURS AND FINAL ORE GRADES

The first analysis of this sort that we know of was done at the Center for Naval Analyses (CNA) in 1984. It was largely based on the performance of 88 carrier-based Navy squadrons in Operational Readiness Evaluations (OREs) between 1980 and 1984 [4], [5]. Although the OREs were given to entire air wings, performance was judged by squadron. The CNA analysis studied the performance of fighter and attack squadrons -- squadrons of F-4s, F-14s, A-6s and A-7s.

ORE information was gathered from both the Atlantic and Pacific fleets. Overall performance in OREs was established via a qualitative grade. These grades were outstanding (the highest grade), low outstanding, high excellent and excellent (the lowest observed grade). This taxonomy is somewhat misleading. Grades of excellent were considered to reflect badly on a squadron. Atlantic Fleet staff members felt that OREs were the best indicator of a squadron's proficiency at the time of the exercise. Pacific Fleet staff members were somewhat less effusive, but still supported their use for analytic purposes.

OREs were tests of operational performance that were graded by observers from outside the airwing (they are no longer given today). Fleet staff members believed that a squadron's overall grade might be less objective than some other specific pieces of information about the evaluations, such as carrier landing grades and boarding rates, which

were felt to be particularly objective. All three of these performance indicators were used in the CNA analysis with consistent results.

Table 1 compares average monthly flying hours for the sampled squadrons in the five or six months before the evaluation with their ORE scores.

Table 1. ORE Grades and Monthly Flying Hours

ORE grade	Average monthly flying hours per squadron
Outstanding	487
Low outstanding	421
High excellent	384
Excellent	356

The correlation between average flying hours and the evaluation grade is clear. A statistical analysis of these data implied that a ten percent decrease in flying hours would result in a 34 percent decrease in the number of squadrons rated outstanding.

Final ORE grades had the virtue of reflecting the totality of squadron performance during the evaluation. They were meant to measure overall proficiency. They were, however, imprecise. It is impossible, for example, to know how much worse high excellent is than low outstanding. We also do not know the degree to which scoring differed among graders. Fortunately, final grades are not the only performance indicator that were saved after OREs. Analyses of both boarding rates and landing grades observed during the OREs reinforce the analysis of final grades.

B. RECENT FLYING HOURS AND CARRIER BOARDING RATES

Table 2 shows the result of deriving a linear relationship between flying hours in the months prior to the ORE and the carrier boarding rate during the ORE. The boarding rate is the fraction of attempted arrested landing passes that are successful. Unsuccessful attempts require an additional pass.

TABLE 2. Equation for Predicting a Squadron's Boarding Rate During ORE

Factor	Coefficient	t-value
Constant	81.9	
Monthly flying hours	0.022*	3.5
$R^2 = .12$	number of observations = 88	

* significant at the 1% level

Another way of writing the equation depicted in Table 2 is: Boarding rate = 81.9 + .022 x monthly flying hours. An implication of this equation is that a ten percent decrease in flying was estimated to yield a ten percent increase in unsuccessful landings.[1] While only a small fraction of the variance in boarding rates was explained, flying hours were highly significant. This means that we cannot do a great job of predicting the boarding rate of any particular squadron, but we can be very confident that if flying hours are cut squadrons in general will experience more unsuccessful landings.

C. RECENT FLYING HOURS AND LANDING GRADES

Every carrier landing is graded by the Landing Signal Officer (LSO) on a four point scale (from 1, the lowest grade, to 4). An analysis of landing grades received during the OREs yielded even more statistically significant, if perhaps less quantitatively important, results. Table 3 shows the results of this analysis.

Table 3. Equation for Predicting a Squadron's Average Landing Grade During ORE

Factor	Coefficient	t-value
Constant	2.83	
Monthly flying hours	.0012	5.6*
$R^2 = .27$	number of observations = 88	

* significant at the 1% level

---

[1] The number of unssuccessful landings would increase from ten percent of total landings to eleven percent of total landings. This is a ten percent increase in the number of unsuccessful landings.

The results imply that a ten percent cut in flying hours would have reduced average landing grades from 3.33 to 3.28 in the squadrons that underwent the OREs in the sample. This would have dropped a squadron with median average landing proficiency to the 38th percentile of the squadrons analyzed in the study.

D. RECENT FLYING HOURS AND BOMBING ACCURACY

The final analysis performed as part of the initial CNA work was not based on OREs. Rather it examined the effect of land-based preparation prior to carrier deployment for an A-6 squadron between February and October of 1983. The indicator of performance was how close to the target the aircraft dropped their bombs. Four kinds of bombing runs were included in the analysis. Over 2500 bombing runs went into producing the data. Since flying-hour information was only available for the entire squadron on a monthly basis, the statistical work was done on an aggregate monthly basis. For each of the four kinds of bombing runs the average miss distance was calculated for every month. To put the four kinds of runs on a comparable basis, each monthly observation was normalized by dividing it by the grand average for that kind of run. Thus, there were 36 monthly observations of normalized bombing accuracy, accuracy relative to average accuracy for the same kind of run.

The examined hypothesis was that practice bombing improves bombing performance (lowers the miss distance). The researchers were able to distinguish time that could have been used to practice bombing (which occurred at Whidbey Island Naval Air Station) from other flying activity (which occurred elsewhere). Table 4 shows the results of looking at normalized bombing accuracy as a function of the amount of flying done at Whidbey Island in the previous month.

The expected relationship holds. Quantitatively, it means that a ten percent cut in flying would increase the average miss distance by 5.2 percent.

9

TABLE 4. Equation for Predicting Normalized Bombing Accuracy

Factor	Coefficient	t-value
Constant	1.51	
Last month's flying at Whidbey Island	-.0018	2.3[*]

$R^2 = .15$ number of observations = 36

* significant at the 5% level

E. TOTAL PILOT EXPERIENCE, BOMBING ACCURACY AND LANDING GRADES

The work cited so far looks at some measure of proficiency as a function of recent flying experience. This is the essential element postulated in the development of flying hour programs, but it isn't the only mechanism likely to be at work. As pilots accumulate flying hours over the course of their careers, they are likely to get more proficient independent of their recent flying experience. This cumulative effect was investigated in a recent Navy study that examined the performance of A-7 pilots in the Western Pacific and at Naval Air Station, Fallon, Nevada. The study found total flying hours to have a significant and substantial effect on both bombing accuracy and landing grades [6]. The analysis did not include individuals with less than 300 hours of experience in jets.

As Table 5 shows, between 300 hours and 2400 hours, a doubling of experience is associated with about 13 percent greater bombing accuracy and with landing grades about 15 percent closer to a grade of 4. Little improvement was discernible above 2400 hours.

TABLE 5. Career Flying Experience, Bombing Accuracy and Landing Grades

Career flying hours in jets	Expected miss distance (feet)[*]	Expected landing grade[**]
300	109	2.96
600	95	3.10
1200	82	3.22
2400	71	3.33

[*] Based on an equation estimated using 208 observations, with $R^2=.39$, and a coefficient significant at the 1% level

[**] Based on an equation estimated using 180 observations, with $R^2=.26$, and a coefficient significant at the 1% level

F. TOTAL PILOT EXPERIENCE AND ACCIDENT RATES

The Naval Safety Center has studied accident rates for Navy tactical aircraft as a function of accumulated pilot experience [7]. Table 6 summarizes the results of this work.

TABLE 6. Safety and Experience in Navy Tactical Aviation -- Mishaps Per 100,000 Flight Hours, 1977-1985

-----------------------------------Hours in Model-----------------------------------			
Under 300	301-500	501-1000	Over 1000
6.52	4.02	3.69	2.49

A statistically significant correlation was found. Pilots who had flown under 300 hours in a particular model of aircraft were about 2.6 times as likely to have an accident as pilots with over 1000 hours of experience.

A factor that complicates interpretation of the results displayed in Tables 5 and 6 is the probability that causality runs both ways. Not only are more experienced pilots likely to be better pilots because they master the necessary skills; intrinsically better pilots are more likely to continue flying long enough to become experienced pilots. This latter relationship pertains both because people are more likely to stick with a job they are

11

particularly good at and because pilots with an especially strong aptitude for flying are less likely to crash early in their careers. Nonetheless, the effect of experience on skill is probably the principal mechanism behind the results reported here. The tables show considerable improvement in proficiency by the time 600 hours of experience is reached. Pilots typically reach this level of experience before they have a chance to leave the service. Accidents are not prevalent enough to have a marked impact on the correlation between skill and experience in the pilot population.

G. BOMBING ACCURACY RELATED TO TOTAL PILOT EXPERIENCE AND RECENT FLYING HOURS

The most detailed attempt to model the relationship between flying history and pilot proficiency was undertaken in a recent Air Force study [8]. Bombing accuracy was studied for pilots of both the F-16 and the A-10. Alone among the studies we have reviewed, this one tried to relate proficiency to both the recent and long-term flying experience of pilots. A complex model of skill growth and deterioration was used to depict the impact of recent flying experience on bombing accuracy. The researchers found that it was best to apply this model separately to pilots with high career experience (over 900 hours in the F-16 and over 1400 hours in the A-10) and lower career experience. This was because of observed correlation between accumulated experience and bombing accuracy.

Figure 2 shows the implications of this analysis. The relative bombing effectiveness scale on the Y-axes in the figure is proportional to the reciprocal of the square of the predicted bombing accuracy of a squadron. Most of the benefits of increased flying displayed in the figure are the result of long-term skill accumulation. Critical plateaus in experience were found for both types of aircraft. More extensive flying-hour programs allow a larger fraction of pilots to reach the higher plateau. Recent flying experience was generally found to have a small impact on bombing proficiency (except for experienced A-10 pilots, whose skills did appear to get honed by practice).

H. PRELIMINARY FINDINGS AND POSSIBLE FUTURE RESEARCH

The literature we have reviewed concentrates heavily or fighter and attack aircraft. It shows that for these types of aircraft the performance of aircrews can be linked to their flying-hour histories. Aircrew proficiency in the Navy has been linked (in separate studies) to both recent flying intensity and accumulated flying experience. Safety in the Navy has been linked to accumulated flying experience. Proficiency in the Air Force has been largely

12

1.0

0.8

RELATIVE
BOMBING   0.6
EFFECTIVENESS

0.4

0.2

0

0   10   20   30   40   50

AVERAGE FLYING HOURS PER PILOT PER MONTH

A-10 Squardron Bombing Effectiveness as a Function of Flying Activity

1.0

0.8

RELATIVE   0.6
BOMBING
EFFECTIVENESS

0.4

0.2

0

0   10   20   30   40   50

AVERAGE FLYING HOURS PER PILOT PER MONTH
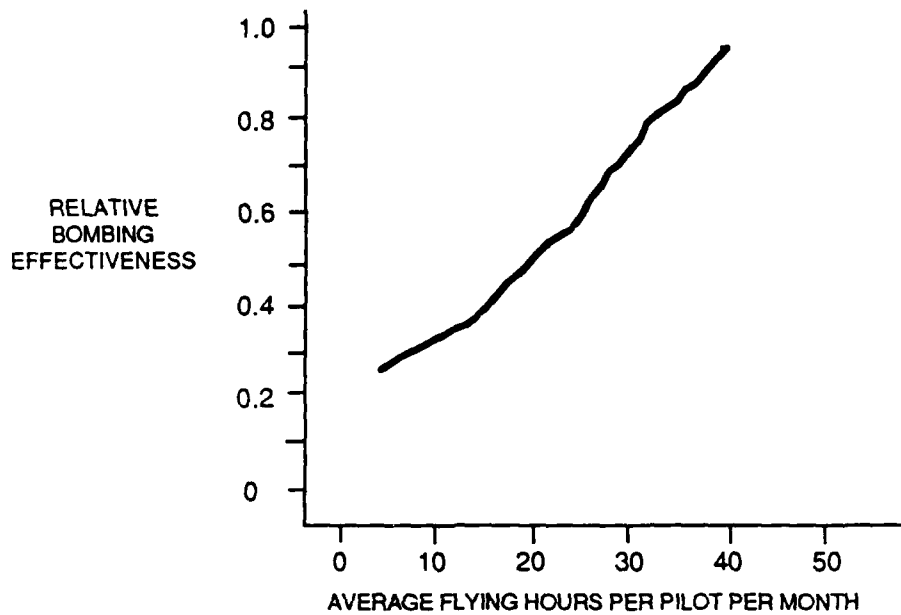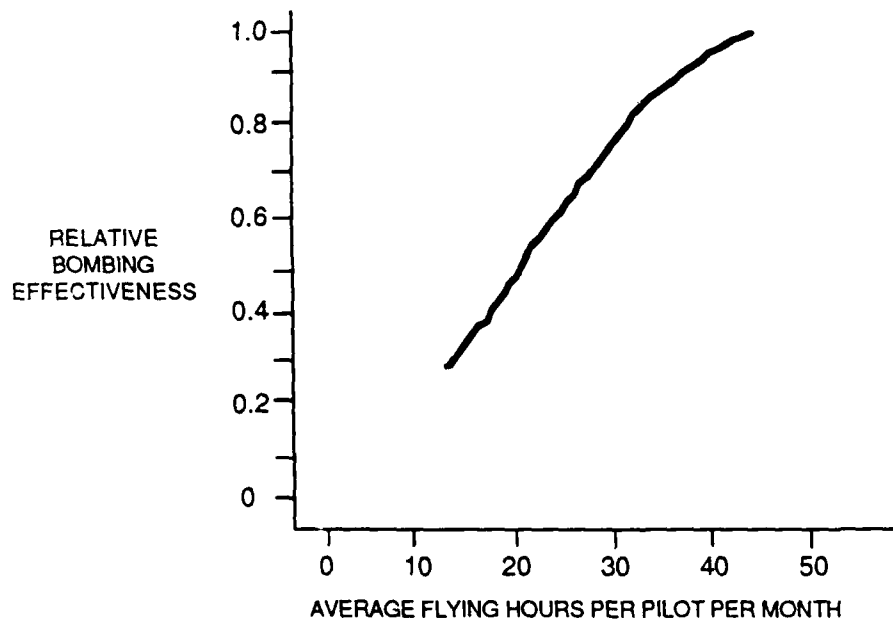
F-16 Squadron Bombing Effectiveness as a Function of Flying Activity

FIGURE 2.   RESULTS OF AIR FORCE ANALYSIS OF BOMBING ACCURACY

13

linked to accumulated flying experience. The kind of statistical analysis we are undertaking clearly can be fruitful.

Still, there are many kinds of aircraft and many missions that have not been analyzed in this way. These include helicopters (we do not know of any Army or Marine Corps analyses that have been performed), transport aircraft, strategic bombers, and Air Force fighters in their air-to-air role (the ORE work included Navy fighters, but did not separate them from attack aircraft). A goal of research in this area should be to extend statistical analysis of the value of training to aircraft types and missions that have not yet been subject to it.

We now turn to a brief discussion of the framework we plan to use to analyze relationships between flying hours and aircrew performance. This will prepare the way for a description of the data on both performance and training histories that we have uncovered in our investigations, and for consideration of the suitability of the data for the kind of analysis we contemplate.

IV. MODELING THE RELATIONSHIPS BETWEEN FLYING HOURS AND AIRCREW PROFICIENCY

A. CONCEPTUAL MODEL

As a result of our literature review, we will adopt a model in which the experience gained through flying more hours manifests itself in two ways: (1) a short-term refreshing of skills that erode without practice, but that can be fairly easily relearned and (2) long-term mastery effects from the incremental increase of total experience over a long period of time. At the present time, only the first of these mechanisms is used to build and justify the flying-hour programs of the services.

Of the studies we have reviewed, only the Air Force study by Cedel and Fuchs examined both effects [8]. Separate relationships between recent flying experience and bombing accuracy were estimated for experienced and inexperienced personnel, with the upper and lower limits of pilot capability modeled as a function of pilot experience. For each group, bombing score was predicted as a function of the time between bombing flights. Unfortunately, they were not successful in finding as pronounced a short-run effect as other researchers have found. This may have been due to the specific model of short-term benefit they used.

We expect to quantify both the short- and long-run effects using multiple regression. Performance is hypothesized to depend on such factors as accumulated flying time, the number of events in a given time period and the elapsed time between events. We also hypothesize an interactive effect between total and recent experience, since total experience may affect how quickly skills are honed and how quickly they decay. Cedel and Fuchs found some support for this latter hypothesis.

If our approach is successful, it should be possible to establish both short- and long-run criteria for flight hour programs. Flying-hour programs would then be oriented to assuring not only that short-run qualification standards are met, but also that a specified fraction of pilots surpass target levels of accumulated experience. A crew member's ability

15

to perform the required mission on call depends on his capability when he is called. Capability when called (readiness), according to the above hypotheses, depends on recent experience and total experience. If the hypotheses are confirmed, each should be a factor in determining the flying-hour program..

Determination to achieve a more experienced mix of pilots is likely not only to enhance combat capability, but to serve the additional purpose of reducing recruiting and initial training costs.

B. STATISTICAL MODEL

We hypothesize that the effects of short-run variables, such as days since last practice or number of flight hours in the last time period (week, month, two months, etc.) will depend on the level of experience the individual aircrewman has. We will test this hypothesis using a functional form which allows us to estimate the effect of interactions.

The basic model is:

$$y = a_0 + a_1 EXP + a_2 X + a_3 (X)(EXP) + u,$$

where

y = Performance measure, such as bombing accuracy, carrier landing grades, etc.

EXP = Experience, such as total flight hours, total time in type, total time in model, etc. A second experience variable, reflecting experience in simulators will be added in some formulations.

X = Short-run variable, such as time since last practice, flight hours or practice flights in the previous week/month/six months, etc. In some formulations X will be a vector of short-run variables. This will allow examination of the possibility that different kinds of recent training (e.g., training in simulators) affect proficiency differently. In these cases a_2 and a_3 will also be vectors.

a_0 = constant

a_1, a_2, a_3 are coefficients and u is an error term.

Different versions of this equation will be estimated using various functional forms (such as linear, logarithmic and logit) depending on the data. Mairs et al. found the distribution of A-7 bomb accuracy to be approximately log normal. In log form, the above equation is a generalized translog production function [9]. The term that includes both the

short-run and experience variables (sometimes called the interaction term) provides information about how the importance of one factor varies with the level of the other.

This analytic approach offers the potential for investigating the effect of competition for resources between different missions. Ultimately we hope to look at proficiency in the performance of individual missions as a function of time spent practicing that mission and time spent practicing other missions. This will allow us to address a major problem expressed by many operators: trying to maintain some minimum level of proficiency in all required missions with limited resources.

V. DATA AVAILABILITY AND SUITABILITY

To demonstrate the feasibility of developing statistical relationships between flying hours and performance for a range of aircraft types, we need data on aircrew performance and flying hour histories for the same aircrews. A wide variety of such data exist. This section discusses our investigations and the data we found. Existing performance data are described in terms of their availability, relevance and reliability. Information on aircrew performance is usually not kept at a central location. If it exists, it tends to exist in the field. Our investigations into the availability of data have taken us to SAC headquarters at Offut AFB outside of Omaha, to Little Rock AFB, where C-130 training is done, to the Navy Safety Center in Norfolk, to the U.S.S. Eisenhower at sea, to the Army Aviation Center at Fort Rucker, Alabama, and to an Army Combat Aviation Brigade at Fort Carson, Colorado. In addition, Washington offices of all four services have been visited. TAC headquarters at Langley AFB was visited as part of an earlier, related study effort.

These visits have identified a great deal of apparently available data. In selecting performance measures to be examined in this study, two principal criteria were used: relevance and reliability. Relevance means that, when properly measured, the variable reflects mission-related performance. Reliability means that there is good reason to believe that measurements are being made in an accurate, reproducible fashion. Some measures, such as bombing scores from an instrumented range, are clearly reliable. Reliability is related to objectivity, the absence of subjectively based variations in measurement, but, the presence of human judgment in the grading process does not necessarily imply a lack of reliability. In cases in which human judgment is present, one should also evaluate the importance attached to the grading process, the degree of standardization of the grading criteria, and the potential consequences of not following the criteria. For example, carrier landing grades are assigned by a Landing Signal Officer (LSO) and depend, in large part, on his judgment. But individual carrier landing performance is one of the most closely tracked records in naval aviation. LSOs are highly skilled and closely monitored by the Air Wing LSO, and landing grades are assigned according to well-defined standards. Finally, individuals above LSOs in the chain of command have a strong interest in both safe

landings and accurate reporting. When an accident occurs, any evidence of laxness in assigning landing grades is viewed as a serious offense.

Validation of performance measures, both the determination of relevance and the determination of reliability, must be based on experience to a great extent and generally requires personal contact with the people who use the measures operationally. Much of this preliminary work in this area was done as part of the earlier research by Hammon and Horowitz [3]. Building on this earlier work, we have assembled candidate performance measures for all the services. With one exception, we have high confidence in their validity.

Data on these performance measures are either in hand or have been promised:

1. Marine Corps air-to-ground scores for fighter and attack squadrons
2. Navy carrier boarding rates
3. Navy bombing scores
4. Fleet Fighter Air Combat Maneuvering Range Program (FFARP) data
5. Navy and Marine Corps mishaps (accidents)
6. Navy Air Effectiveness Measurement (AIREM) performance
7. Air Force bomb and missile scores
8. Air Force mishap rates
9. Air Force C-130 drop scores
10. Navy carrier landing grades
11. Navy and Marine Corps flight check (NATOPS) grades
12. Air Force Standardization/Evaluation check flight scores
13. Army Standardization Flight Evaluation results

These performance measures all meet the relevance criterion. If properly measured, they have a clear link to effective mission performance. Most are self-explanatory, but a few are not. AIREM data are the result of realistic anti-submarine warfare (ASW) exercises, including weapons drops. Success in identifying and killing target submarines is noted. The flight check data for all the services reflect both knowledge of procedures and execution of flight and tactical maneuvers.

Most of the data are also clearly reliable. The bombing and missile scores are derived from physical or electronic measurement by outside observers. The Navy Air

Combat Maneuvering Range (ACMR) data are developed by electronic means on an instrumented range.[2] So are AIREM data. Navy boarding rates are determined by direct computation based on whether or not the pilot successfully completes an intended full-stop landing. Carrier landing grades incorporate some degree of subjectivity, but, as was noted above, they are characterized by a high degree of standardization and the risk of adverse consequences for inflated grading.

Flight checks are graded by certified examiners. In the Air Force, Navy and Marine Corps, care is taken to insure as much objectivity as possible. Evaluation content and grading criteria are standardized in detail by aircraft model and series. In most cases specified procedures leave little room for subjectivity by the evaluator. Most criteria are stated in quantitative terms, such as how much variation from a desired altitude level is allowed. If altitude varies less than that amount, the grade for the maneuver is a pass; otherwise it is a fail. Each model is managed by a Type Commander or Major Command and the results are taken very seriously by the services. Research into the variation of grades over evaluators is currently being conducted by the Air Force Human Resources Laboratory. We are inclined to believe that Air Force, Navy and Marine Corps flight check grades are reliable enough to be included in our empirical analyses, but the on-going research should be followed for additional evidence on this point.

While the Army relies on its flight checks as the other services do, the evaluation criteria appear to be specified in somewhat less detail, raising the risk of subjectivity and unreliability. Unfortunately, flight evaluations are the only performance measure we have been able to gather for Army helicopter crewmen. Any quantitative work that is based on these data should be treated as exploratory.

Turning to data on flying-hour activity, the Air Force, Navy and Marine Corps all have centrally available, automated information on the flying-hour histories of individual aircrew members. The Air Force Operations Resource Management System (AFORMS) is a standardized reporting and data base system for training information. Current experience and training data are maintained for all active duty personnel. Information includes experience (total and by aircraft type), combat time, and monthly flight time and sorties during the current six month period. AFORMS is kept at the Major Command level. It

---

[2] The Navy refers to the monitoring equipment on such a range as a Tactical Air Combat Training System (TACTS), the Air Force refers to it as Air Combat Maneuvering Instrumentation (ACMI).

feeds the HORIS (Hormats' Information System) data base which is maintained at Air Force Headquarters. HORIS keeps monthly data for a year and annual data before that.

In January 1987 the Navy began to use the Naval Flight Information Reporting System (NAVFLIRS). It includes data on flying hours and simulator usage. NAVFLIRS data must be submitted after every flight. Before the institution of NAVFLIRS, the Navy used the Individual Flight Activity Reporting System (IFARS). NAVFLIRS and IFARS data are summarized by month and fiscal year. Validated IFARS data are available from the Naval Safety Center.

The Marine Corps also uses NAVFLIRS. Its Flight Readiness and Data System (FREDS) was the forerunner of NAVFLIRS. In January 1987, the Marines began to report ordnance delivery performance to NAVFLIRS.

The Army does not have centralized information on the flying-hour histories of its aircrewmen. Hard copy records are kept at the brigade level.

VI. PLANNED AND PROSPECTIVE ANALYSES

Suitable data appear to be available to perform many different analyses of quantitative relationships between flying hours and relevant and reliable measures of aircrew performance or safety. In view of the limited resources available for phase two of this study, choices must be made about which analyses to perform and the order in which to perform them. These choices will be made according to three criteria: the speed with which data are acquired, the desire to produce analyses covering as wide a range of services and aircraft types as possible, and policy interest in particular services, aircraft types or measures of performance. An example of policy interest is the desirability of addressing the GAO's comments on the supportability of the flying-hour programs for TAC and SAC aircraft.

We want to develop as many quantitative relationships as possible. In the first phase of this project, requests have been made for data from many sources. Under such circumstances, researchers cannot always predict what data they will be able to acquire first. The data that we were able to acquire first are not necessarily the data we most want to analyze, but the way to expedite the development of quantitative relationships is to analyze data as they become available, rather than let acquired data sit unanalyzed while effort is focused on acquiring additional data. Decisions concerning what data sets to try to develop next will be made according to the second and third criteria.

We plan to start our empirical work by analyzing two data sets that have already been assembled. These initial studies will examine the impact of variations in flying hours on the accuracy with which Marine Corps aviators deliver air-to-ground ordnance and on various measures of performance for Navy tactical aviation. The remainder of this section begins by providing an overview of the initial studies. Our preferences about what additional studies to pursue are then discussed.

A. MARINE CORPS ANALYSIS

The Marine Corps data set is the richest, in detail and number of observations, of those in hand. Performance measures, however, are limited to air-to-ground ordnance delivery. In January 1987 the Marine Corps began entering air-to-ground accuracy for all

flights for which an outside observer was present. In most cases the outside observer is located at an instrumented range. The data set includes information on performance, short-run experience and total experience.

Performance information is recorded for flights flown by approximately 90% of fighter/attack squadrons for the period January through September of 1987. This file includes nearly all flight data recorded on the NAVFLIRS flight log form (yellow sheet). This includes flight purpose and training codes, flight hour, landing and approach information, and bombing accuracy by type of delivery.

Information on recent experience covers all pilots in the performance data base for June 1986 through September 1987. This data base is also by flight, and includes essentially all flight information in the performance data base except performance (bombing accuracy). The data base covers all pilots and Naval Flight Officers (NFOs) who appear in the performance data base. Data include all flight experience, including the use of simulators and experience as a student, for the six-month period. Since flight purpose and training codes are included, detailed short-run experience variables can be constructed for the six-month period preceding the period observed in the performance data base.

Total experience is measured as of June 30, 1986. The data base includes the year in which an individual was designated an Aviator or NFO, number of months assigned to an operational squadron, total flight time (day and night) and total flight/night hours, landings and approaches in type and model, and simulator time.

B. NAVY CARRIER AIRWING ANALYSIS

This data base includes carrier landing grades, boarding rates, ACM scores and bombing accuracy for an Atlantic Fleet carrier airwing. Eight squadrons (two fighter, three attack, one electronic countermeasures, one air antisubmarine, one early warning) are represented. The period covered is August 1986 through October 1987. Landing information is by date, and includes data for seven at-sea periods. The last four at-sea periods constituted the airwing's work-up for a major deployment and Advanced Phase Evaluation (the ungraded successor to the Operational Readiness Evaluation). Individual flight statistics for pilots and Naval Flight Officers include total flight hours and carrier landings, total flight hours and carrier landings in current model, and flight hours and carrier landings by month. All flight statistics are broken down by day and night.

23

Landing grades are taken from the standard trend (grade) sheets. Boarding rates are calculated directly from the trend sheets. Fighter squadron ACM data are extracted from the most recent Fleet Fighter ACM Readiness Program (FFARP). Performance measures include survival time and kill/killed scores. FFARP flights are flown on an instrumented range against instructors who are assigned full time to an ACM training squadron.

Bomb scores are for the most recent Competitive Exercise. Daily scores are available for only one squadron. If time is available, we will collect AIREM data for the ASW aircraft in this same airwing. This would give us nearly full coverage of the primary mission areas for the entire wing.

C. PROSPECTIVE ANALYSES

As the two analyses described above proceed, data will be gathered to permit additional case studies to be performed. The top priority will be placed on trying to extend the work to examine the determinants of performance in Air Force and Army aircraft.

Analysis of Air Force Data. For the Air Force, attempts will be made to set up analyses for bomber, fighter and transport aircraft. The First Combat Evaluation Group at SAC Headquarters has agreed to supply machine-readable information on several indicators of performance for various members of B-52 crews. Heading error and the ability to hold low altitude could be used as indicators of pilot proficiency. Bomb and missile accuracy may depend most on the proficiency of the radar navigator. Jamming effectiveness on electronic warfare (EW) ranges could serve as an indicator of the performance of EW officers.

A decision has not yet been reached about what measures of performance to concentrate on for fighter aircraft. Standardization/Evaluation (STAN/EVAL) results could be used, but analysts at the studies and analysis office at Air Force Headquarters are somewhat leery of them. They note that a very high proportion of pilots pass their STAN/EVALs. Thus, important variations in performance may be missed if they are the sole source of proficiency data. These analysts have suggested following a survey approach. This would involve asking officers to rank all the pilots in their squadron and using the resulting rankings as the measure of proficiency. Designing and implementing such a survey could prove beyond the resources available to us. Another alternative is to rely on data developed by monitoring individual performance at the squadron level. This deserves further investigation.

Analysts being supported at Little Rock AFB by the Air Force Human Resources Laboratory are engaged in research on the performance of C-130 aircrews. They have recommended that STAN/EVAL results and the accuracy with which materiel is air-dropped be used as indicators of performance for the C-130. An attempt will be made to obtain data on these measures. The MAC STAN/EVALs are not graded on a strict pass/fail basis. A moderate fraction of the aircrewmen evaluated receive provisional passes. This yields a data set with more information on gradations in performance.

Analysis of Army Data. The only performance indicator for Army helicopter crews that we have identified is the outcome of individual flight evaluations. As was noted earlier, Army evaluations are probably less objective than those performed by the other services. Nonetheless, we are very interested in determining the feasibility of relating flying hours to measures of performance for all the services. For this reason an exploratory Army case study that uses flight evaluation data seems worthwhile. Army personnel both at Fort Rucker and at Fort Carson have been cooperative about supplying such data. We understand that the Army data may not be good enough to support the planned quantitative analysis. This analysis is best viewed as an investigation into the limits of the feasibility of relating flying-hour histories to aircrew proficiency.

VII. CONCLUSIONS

This paper presents the results of an investigation into the practicability of using a statistical approach to develop quantitative relationships between the prior training received by aircrew personnel and indicators that are clearly related to how well they can be expected to perform in combat. Our conclusions are:

1. It is feasible to relate flying-hour activity to operational performance and safety measures. It has been done. While the research in this area has not been extensive, the published analyses have generally produced results that seem to be credible. Quantitative relationships of the kind we seek have been developed. They support the proposition that more flying results in measurably better performance. This has been demonstrated for both Air Force and Navy aircrews.

2. There is reason to believe that additional flying affects the level of aircrew proficiency in two ways. In the short run it appears to hone skills and prevent their deterioration. In the long run it permits aircrew members to achieve a higher level of mastery that is reflected in better performance. None of the existing analyses that were reviewed fully captured both of these effects. Only one tried. Empirical work should follow an approach that allows both the short-run and long-run effects of variations in flying hours on aircrew proficiency to be quantified.

3. Proficiency data exist for all the aircraft types we have investigated. In most cases they are both relevant to our purpose and clearly objective. In addition, most of the indicators of proficiency that reflect evaluator judgment are developed in a highly structured fashion that seems to preclude much undesirable subjectivity.

4. The services are willing to support efforts to gather the data that are needed to perform statistical analyses.

5. Data exist to develop links between flying-hour activity and measures of operational performance and safety for a wide range of aircraft. Both justification and formation of flying-hour policies would benefit from them. Additional research to build such links should be supported.

26

REFERENCES

[1]   "Aircrew Training: Tactical Air Command and Strategic Air Command Flying Hour Programs," Briefing Report to the Chairman, Subcommittee on Defense, Committee on Appropriations, House of Representatives, United States General Accounting Office, September 1986.

[2]   Henry Solomon, "Economic Issues in Cost-Effectiveness Analyses of Military Skill Training," IDA Paper P-1897, 1986.

[3]   Colin P. Hammon and Stanley A. Horowitz, "Relating Personnel and Training Resources to Unit Performance: Identifying Data on Performance in the Military," IDA Paper P-2023, 1987.

[4]   Linda Cavalluzzo, "OPTEMPO and Training Effectiveness," Professional Paper 427, Center for Analyses, December 1984.

[5]   Stanley A. Horowitz, Linda Cavalluzzo and CDR Gary Johnson, "Flying Hours and Aircrew Performance," CNA Working Paper 1540.10, September 1984.

[6]   Lee Mairs, Russell McKennan, Albert Lo, Cynthia Miller and William Truscott, "Quantifying the Training to Readiness Link: the Case for Attack Pilots," Systems-Analytics Group, November 1986.

[7]   Michael S. Borowsky, "Readiness and Retention: Pilot Flight Experience and Aircraft Mishaps," U.S. Naval Safety Center, Statistics and Mathematics Department, June 1986.

[8]   Lt. Col. Thomas E. Cedel and Lt. Col Ronald P. Fuchs, "An Analysis of Factors Affecting Pilot Proficiency," Fighter Division, Directorate of Theater Force Analyses, Air Force Center of Studies and Analyses, December 1986.

[9]   See R. S. Brown, D. W. Caves and L. R. Christensen, "Modeling the Structure of Cost and Production for Multiproduct Firms," *Southern Economics Journal*, Vol. 46, Number 1, July 1979, pp. 256-273 for a discussion of the translog joint cost function.