AD-A091 229 DAYTON UNIV OH RESEARCH INST F/G 5/9
POP-UP WEAPON-DELIVERY MANEUVER: USE OF PILOT SELF-ASSESSMENT D--ETC(U)
OCT 80 D R LYON, J L EUBANKS, T H KILLION F33615-77-C-0054

UNCLASSIFIED AFHRL-TR-80-33 NL

2

END
DATE
FILMED
12-80
DTIC

CONT.

AFHRL-TR-80-33

LEVEL II

AIR FORCE HUMAN RESOURCES LABORATORY

AD A091229

POP-UP WEAPON-DELIVERY MANEUVER: USE OF PILOT SELF-ASSESSMENT DATA IN ANALYSIS OF CRITICAL COMPONENTS

By

Don R. Lyon
James L. Eubanks
Thomas H. Killion
Robert T. Nullmeyer
University of Dayton Research Institute
300 College Park Avenue
Dayton, Ohio 45469

Edward E. Eddowes

OPERATIONS TRAINING DIVISION
Williams Air Force Base, Arizona 85224

October 1980

Final Report

Approved for public release; distribution unlimited.

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

NOTICE

When U.S. Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This final report was submitted by the University of Dayton Research Institute, 300 College Park Avenue, Dayton, Ohio 45469, under Contract F33615-77-C-0054, Project 1123, with the Operations Training Division, Air Force Human Resources Laboratory (AFSC), Williams Air Force Base, Arizona 85224. Dr. Edward E. Eddowes was the Contract Monitor for the Laboratory.

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

MARTY R. ROCKWAY, Technical Director
Operations Training Division

RONALD W. TERRY, Colonel, USAF
Commander

SUBJECT TO EXPORT CONTROL LAWS

This document contains information for manufacturing or using munitions of war. Export of the information contained herein, or release to foreign nationals within the United States, without first obtaining an export license, is a violation of the International Traffic in Arms Regulations. Such violation is subject to a penalty of up to 2 years imprisonment and a fine of $100,000 under 22 U.S.C. 2778.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFHRL-TR-80-33	2. GOVT ACCESSION NO.: AD-A092 229
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): POP-UP WEAPON-DELIVERY MANEUVER: USE OF PILOT SELF-ASSESSMENT DATA IN ANALYSIS OF CRITICAL COMPONENTS	5. TYPE OF REPORT & PERIOD COVERED: Final
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Don R. Lyon, James L. Eubanks, Thomas H. Killion, Robert T. Nullmeyer, Edward E. Eddowes	8. CONTRACT OR GRANT NUMBER(s): F33615-77-C-0054
9. PERFORMING ORGANIZATION NAME AND ADDRESS: University of Dayton Research Institute, 300 College Park Avenue, Dayton, Ohio 45469	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 62205F; 11230236 and 11230337
11. CONTROLLING OFFICE NAME AND ADDRESS: HQ Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base, Texas 78235	12. REPORT DATE: October 1980
	13. NUMBER OF PAGES: 30
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Operations Training Division, Air Force Human Resources Laboratory, Williams Air Force Base, Arizona 85224	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

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

18. SUPPLEMENTARY NOTES

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

flying training
pilot training
skill measurement
skill maintenance
self-assessment
weapons delivery
task analysis

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

Performance of the pop-up weapon-delivery maneuver by F-4 pilots and Weapons Systems Officers was analyzed by collecting detailed ratings and reported errors on segments of the maneuver. The major result of this analysis was that rated performance on the final few seconds of the maneuver, during which the crew is trying to execute a constant angle, high-speed dive, is clearly the best predictor of weapons delivery accuracy. Training implications and issues related to the use of self-assessment data are briefly discussed.

DD FORM 1 JAN 73 1473

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SUMMARY

F-4 Pilots and Weapons Systems Officers (WSOs) from nine mission ready squadrons across the United States provided rating and error evaluation data on their performance of the pop-up weapon-delivery maneuver. The maneuver was subdivided into eight components. Aircrews assessed their performance on each component of 545 scored maneuvers. They also listed causes of less-than-optimal performance on particular components of each delivery. This information was compiled and analyzed to determine the relative contribution of each of the components to the accuracy of weapon delivery.

The major result of these analyses was that rated performance on the final few seconds of the maneuver, during which the crew is trying to execute a constant angle, high-speed dive, is clearly the best predictor of weapon-delivery accuracy. Several aspects of the data show this finding is robust and not artifactual. The same result was obtained for both pilots and WSOs at each wing and for both ratings and reported error frequencies.

The outcome of this study indicates the analysis of self-assessment data can, under some conditions, yield important information about the components of a skill. The use of self-assessment data, however, is not advocated in situations where reasonably economical and unobtrusive direct measures of performance are available.

PREFACE

This study was conducted by the Operations Training Division, Air Force Human Resources Laboratory, Air Force Systems Command. The study supports Project 1123, Flying Training Development, Mr. James F. Smith, project monitor; Task 112302, Training Methods and Media, Dr. Bernell J. Edwards, task scientist; and 11230236, Project SMART, Dr. Edward E. Eddowes, work unit manager.

The authors extend their appreciation to the pilots and weapon systems officers of the 334th, 335th, and 336th Tactical Fighter Squadrons of the 4th Tactical Fighter Wing, Seymour Johnson AFB, NC; the 68th, 70th, and 339th Tactical Fighter Squadrons of the 374th Tactical Fighter Wing, Moody AFB, GA; and the 428th, 429th, and 430th Tactical Fighter Squadrons of the 474th Tactical Fighter Wing, Nellis AFB, NV who provided the information on tactical fighter pilot performance on which this study was based.

TABLE OF CONTENTS

LIST OF FIGURES

POP-UP WEAPON-DELIVERY MANEUVER: USE OF PILOT DATA IN ANALYSIS OF CRITICAL COMPONENTS

INTRODUCTION

This report is based on data collected during a large scale test of methodology developed for the Air Force Skill Maintenance and Reacquisition Training Program (Project SMART), which has the goal of defining and measuring the basic skills which support aircrew mission readiness. The test was requested by Tactical Air Command Headquarters during a preliminary evaluation of Project SMART. The purposes of the report are twofold: (a) present the results of a self-assessment methodology explored by Project SMART to analyze skill in pop-up weapon delivery and (b) document some of the strengths and weaknesses found with this approach. The present report is concerned with the results and effectiveness of the methodology and a detailed description of preliminary development efforts presented in Pierce, DeMaio, Eddowes, and Yates (1979).

In order to adequately assess mission readiness, Project SMART has focused its data collection efforts on continuation training. Initial efforts consisted of interviewing pilots at operational squadrons to determine (a) the tasks most critical to their mission and (b) the important parameters underlying performance of these tasks. On the basis of the interviews, the pop-up weapon delivery (Pierce et al., 1979) and low altitude tactical formation (DeMaio, Eddowes, 1980) tasks were selected for further study. Aircrews were then asked to make detailed assessments of their own performance on these tasks. This report concerns the analysis of pop-up weapon delivery skill based on these data.

Collecting subjective data from experts might be expected to have some inherent advantages. It should have no impact on operational equipment and does not require long data collection sessions that might interfere with scheduling. Because of this, it is sometimes considered an acceptable last resort in situations where actual performance data cannot be collected. Unfortunately, aircrew self-assessment may also be viewed as a potentially misleading source of data. Nisbett and Wilson (1977) reviewed a number of studies that demonstrate the difficulty of making correct inferences from subjective reports, even when such reports do not necessarily imply self-evaluations.

Can pilot self-assessment be a useful source of data for understanding the components of mission readiness? Before the specific case examined in this report is discussed, the general issues involved will be reviewed a little more closely.

USE OF PILOT SELF-ASSESSMENT

Consider the following common situation. A pilot performs some maneuver that can be objectively scored (for example, navigation to a point, or weapon delivery). The pilot then is informed of the score received. After returning from the sortie, the pilot may be asked for an evaluation of how well the maneuver was performed. There are several ways in which the pilot may arrive at this evaluation. At best, the pilot might try to recall the particular characteristics of the performance and compare them to some performance ideal. However, the pilot might instead give the answer implied by the score without considering other aspects of the performance. Or worse, the pilot could simply give the evaluation believed to be the most pragmatic, without referring to the actual performance at all.

One way to distinguish between these possibilities is to ask for evaluations of the components of the performance instead of just an overall evaluation. If the pilot is referring only to information about the desirability of the evaluation or the known success of the maneuver, then the evaluation of all the components should covary strongly, or receive roughly the same evaluation. However, to the extent that the evaluation of different task components shows different characteristics, confidence increases that some component-specific knowledge is being tapped; that the pilot is making an effort to accurately assess the performance.

Unfortunately, even if different components of performance receive different ratings, this does not prove that the pilot's judgments are based on accurate memory for performance. Component ratings could also be based on reconstructions of plausible performance. For example, if it is part of general pilot lore or accumulated personal experience that a particular part of a maneuver is the most difficult, self-evaluation of this part might be adjusted accordingly.

Fortunately, there are some non-experimental techniques for distinguishing between memory for actual performance and plausible reconstruction. One way is to ask the pilot to recall aircraft parameters (airspeed, altitude, etc.) which characterized each component. Ideally, these could be compared to the parameters that the pilot was trying to achieve. If consistent differences exist, then evidence is gained supporting accurate memory for performance.

Another approach, involving still more detail, is to require the pilot to recall the specific errors made. The lack of reported errors will not be very revealing, but their presence provides data to compare with low self-evaluations. As a general principle, one should require as much detail in self-assessment as the constraints of the situation will permit.

In the study described here, both component ratings and reported error data were collected on performance of a short duration, relatively difficult maneuver (pop-up weapon delivery).

TASK DESCRIPTION

The pop-up weapon delivery was identified by pilots very early in the Project SMART effort as a critical air-to-ground maneuver in the tactical fighter pilot's repertoire. The maneuver is designed to be used after the pilot has flown a low altitude route to a target. After reaching a preplanned pull-up point, the pilot climbs quickly and rolls over to an apex altitude and then dives from a prespecified position (called the track point) at a constant dive angle toward the release point, from which a weapon is released. For purposes of analysis, this maneuver was broken down into meaningful discrete segments on the basis of interviews with pilots (Pierce et al., 1979). Figure 1, taken from Pierce et al., shows the resulting segmentation of the maneuver.

In the present study, mission-ready F-4 pilots were asked to make detailed assessments of the crew's performance on the aforementioned components of the pop-up delivery. An examination of the relationships between component assessments and accuracy was expected to isolate the critical components on the maneuver.

METHOD

Participants

Pilots and WSOs from nine operational F-4 squadrons participated in the study. Squadrons participating were those of the 474th Tactical Fighter Wing (TFW) at Nellis AFB, the 347th TFW of Moody AFB, and the 4th TFW at Seymour Johnson AFB.

Procedure

Pilots and WSOs were asked to make detailed assessments of the crew's performance on each of the aforementioned segments of the pop-up each time the maneuver was performed. These assessments took the form of ratings on a four-point scale (excellent, satisfactory, marginal, and unsatisfactory) of performance on each segment and written comments consisting largely of explanations of errors made on each segment. In addition, bomb scores (miss distances) were recorded for each controlled range delivery, and the delivery outcome (bull, hit, miss, dry, or abort) was recorded for tactical range deliveries.

These assessments were gathered on the form shown in Figure 2 according to the detailed instructions (Figure 3). During the debriefing session of daily sorties, approximately 80 percent of these forms were collected by members of the Project SMART staff, who were present to answer questions about the procedure. The remainder of the forms were filled out in the absence of a project researcher by crews who had nevertheless been formally briefed on the assessment procedure.

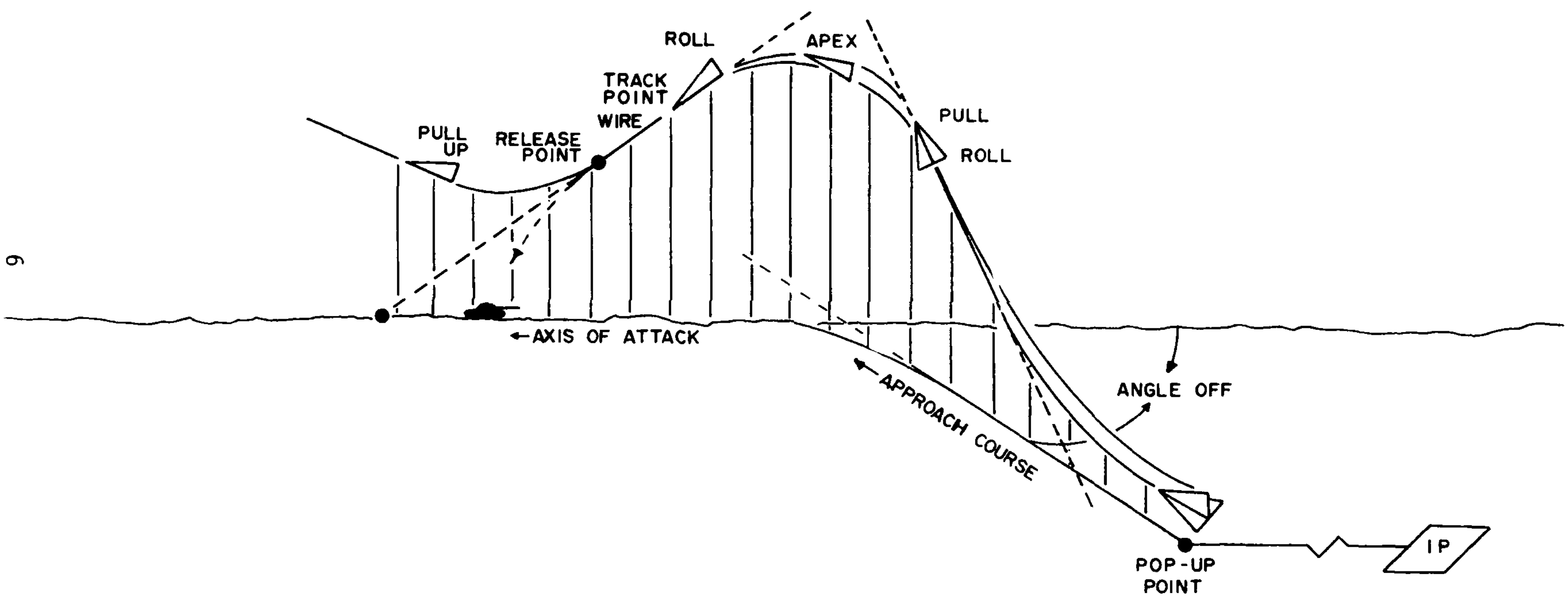

FIG. 1. POP-UP WEAPON DELIVERY PROFILE (From Pierce et al., 1979).

POP-UP EVALUATION FORM

PILOT #:__________ SQUADRON #: ________ RANGE #:__________

EVENT:__________ BLOCK:__________ DATE:__________

	PASS #				
	1st	2nd	3rd	4th	
BOMB SCORES	___	___	___	___	
Task Evaluation					COMMENTS/INDICATE PASS #
1. Approach to PUP	___	___	___	___	__________
2. PUP	___	___	___	___	__________
3. Climb Leg	___	___	___	___	__________
4. Target Acquisition	___	___	___	___	__________
5. Pull Down Point	___	___	___	___	__________
6. Apex	___	___	___	___	__________
7. Track Point	___	___	___	___	__________
8. Bomb Run	___	___	___	___	__________
9. Recovery	___	___	___	___	__________
10. Rtn to Low Alt	___	___	___	___	__________
11. Exposure Time	___	___	___	___	__________

Legend: E - Excellent S - Satisfactory M - Marginal U - Unsatisfactory

FIG. 2. SELF-ASSESSMENT FORM FOR POP-UP WEAPON DELIVERY.

INSTRUCTIONS

1. COMPLETE THE PILOT IDENTIFICATION PORTION OF THE FORM

2. RECORD BOMB SCORES.

3. GRADE THE TASK EVALUATION SECTION AS FOLLOWS:

   E - Excellent. Task performance met criteria with no error reflecting an unusually high degree of ability; no compensations were required.

   S - Satisfactory. Task performance met criteria with minimal error; minimal compensations were required.

   M - Marginal. Task performance met criteria with error; compensations were required to salvage the pass delivery.

   U - Unsatisfactory. Task performance did not meet criteria; gross errors in performance led to either an unsafe or aborted pass.

Task evaluations are to be based on 1) proficiency to maneuver the aircraft, 2) situation awareness, 3) agressiveness and 4) survivability. Any item graded as either M or U requires an appropriate explanation under the Comments section.

The following indicates those requirements identified with each item included in the Task Evaluation section:

1. Approach to PUP: (a) Acquisition of PUP; (b) Altitude control; (c) Airspeed control; and (d) Heading control.

2. PUP: (a) Heading correction; (b) "G" application; (c) Airspeed correction; and (d) timing/distance error.

3. Climb Leg: (a) Climb angle corrections; and (b) airspeed corrections.

4. Target Acquisition: Self-explanatory

5. Pull Down Point: (a) Roll; (b) Airspeed corrections; (c) "G" application; and (d) Altitude/position control.

6. Apex: (a) Pattern/position correction; and (b) Airspeed corrections.

7. Track Point: (a) Aim off point; (b) Roll out; (c) Initial wind correction; (d) Angle, azimuth, and position check; and (e) Initial pipper placement.

8. Bomb Run: (a) Aiming error corrections; (b) Airspeed control; (c) Tracking time control; and (d) Altitude, azimuth, and dive angle corrections.

9. Recovery: (a) "G" application; (b) Jinking; and (c) Altitude and timing control

10. Return to Low Altitude: Transition to low altitude.

11. Exposure Time: Minimization of total time spent out of low altitude environment.

FIG. 3. INSTRUCTIONS ACCOMPANYING SELF-ASSESSMENT FORM.

A very similar self-assessment procedure has been shown to produce segment ratings which, when combined into an overall score for each delivery, are related to bombing accuracy of A-7 pilots (Pierce et al., 1979). However, due to a Tactical Air Command requirement that this research minimize interference with the normal training routine, it was not possible in either the Pierce et al. (1979) study or the current one to withhold from the crewmembers knowledge of their actual bomb scores. Thus, the simplest hypothesis to explain any overall tendency for higher ratings and low bomb scores to be related is that the ratings do not reflect actual performance on individual segments, but rather are based on retrospective inferences from the success or failure of the delivery. This explanation predicts that there should be no systematic differences in the characteristics of the ratings for different segments. However, the present report contains analyses of the segment ratings which show very sharp differences between segments. Such differences reflect the use of segment-specific information in producing segment ratings.

Analysis

To look directly at the extent to which rated performance on each pop-up segment was related to bombing accuracy, the following analysis was performed. First, bomb scores for the controlled range deliveries were categorized as bulls, hits, or misses according to the criteria used for this categorization by the squadrons themselves. These criteria (given in Appendix A) differ for different events and are the same as the criteria by which the tactical range scores are categorized. Such categorization allows scores for tactical and control ranges and scores from different events to be analyzed together, as well as allowing separate analyses to be directly compared. Next, categorized bomb scores and ratings for each of the first eight segments of the pop-up (approach to pull-up point through bomb run) were cross-classified. In order to obtain sufficient cell frequencies for a chi-squared test, bomb score categories were collapsed so that bull and hits were in one category and the misses, dry passes, and aborted runs were in another. Also, ratings were collapsed into two categories with ratings which do not reflect significant error (excellent and satisfactory) in one category and marginal and unsatisfactory ratings in the other. Chi-squared values and values of the chi-squared contingency coefficient (o) were computed on the resulting 2 x 2 matrices for each of the first eight segments of the pop-up (the segments that precede the bomb delivery).

Another source of data on the question of the criticality of pop-up segments is the comments which supplemented the segment ratings on about one-third of the deliveries. Many of these comments mentioned specific errors made on a particular segment, and there were a sufficient number of errors mentioned to allow computation of the proportion of deliveries which resulted in misses or dry or aborted passes, given that a particular kind of error was reported. This proportion indicates the severity of the error. It can be compared with errors introduced on different parts of the maneuver. Errors which were clear consequences of errors made earlier in a delivery were not included in this analysis.

A number of other analyses, conducted to examine more specific issues, are described in the results section. Because of the number of statistical tests conducted, the criteria for significance used are p .0025 for each of four tests of entire distributions and p .001 for each of 37 tests of specific segment/bomb score relationships. These values yield an overall probability (of accepting a chance result as significant) of less than 0.05.

RESULTS AND DISCUSSION

The pilots' ratings of segment performance are considered first. Figure 4 shows the obtained values for the strength of association between segment ratings and bombing accuracy for each segment of the pop-up, based on 545 scorable deliveries from nine operational F-4 squadrons. The overall pattern shows that performance ratings on the final two segments before ordnance delivery, i.e., track point and bomb run, have by far the strongest relationship to bomb accuracy. Ratings by the WSOs for the same deliveries show substantially the same pattern, though the association of ratings with accuracy is somewhat lower. These data are shown in Figure 5. WSOs do not get as clear a view of the maneuver as does the pilot; nor are they directly involved in controlling the aircraft. Therefore, the more detailed analyses reported below were conducted on the pilot data only. Nevertheless, the WSO data confirm that track point and bomb run ratings are the best predictors of accuracy.

Much the same pattern of results holds for maneuvers with both high (30°) and low (10° to 15°) delivery angles. Figures 6 and 7 show the results of a breakdown of the data into high versus low angle deliveries. The figures show that for low angle deliveries only, the predictivity of the target acquisition rating (segment 4) is significant. Unfortunately, on this one segment for this partitioning of the data, there are too few observations in one of the cells of the matrix to ensure that the chi-squared estimate of significance is accurate. However, the main result is clear: track point and bomb run segments were the best predictors of accuracy for both kinds of deliveries.

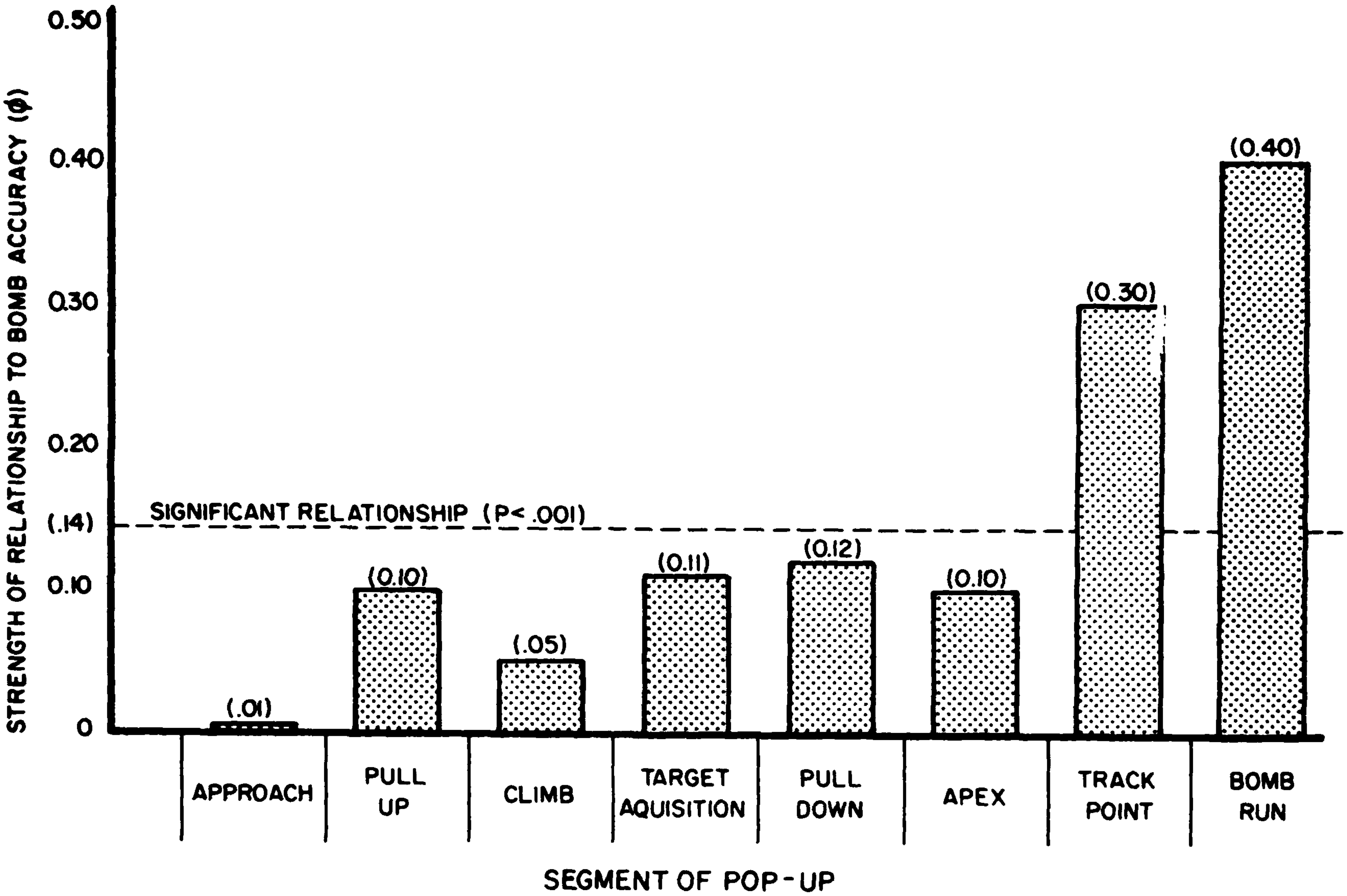

FIG. 4. RELATIONSHIP OF SEGMENT RATINGS TO BOMBING ACCURACY, ALL PILOTS.

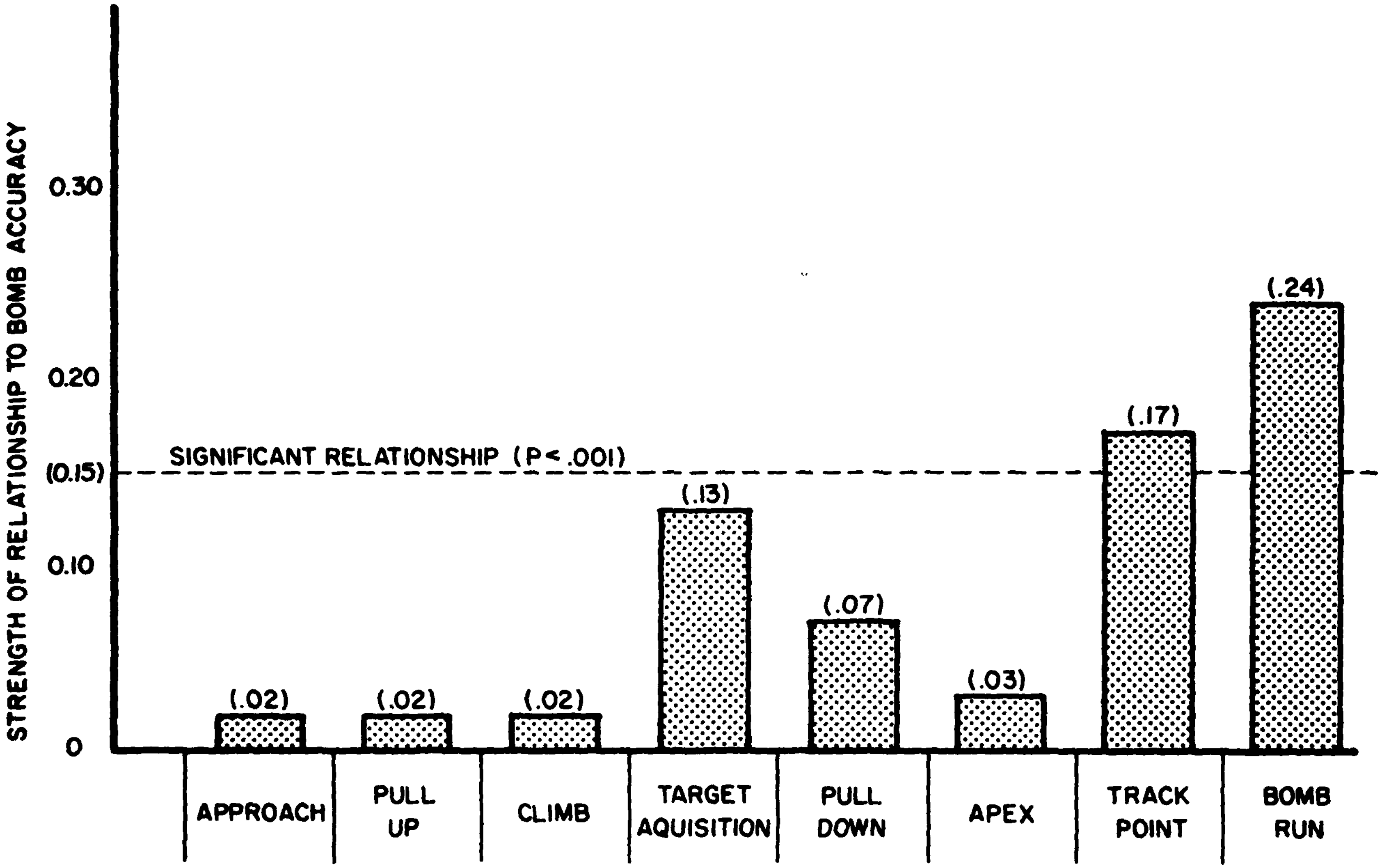

FIG. 5. RELATIONSHIP OF SEGMENT RATINGS TO BOMBING ACCURACY, WSO'S.

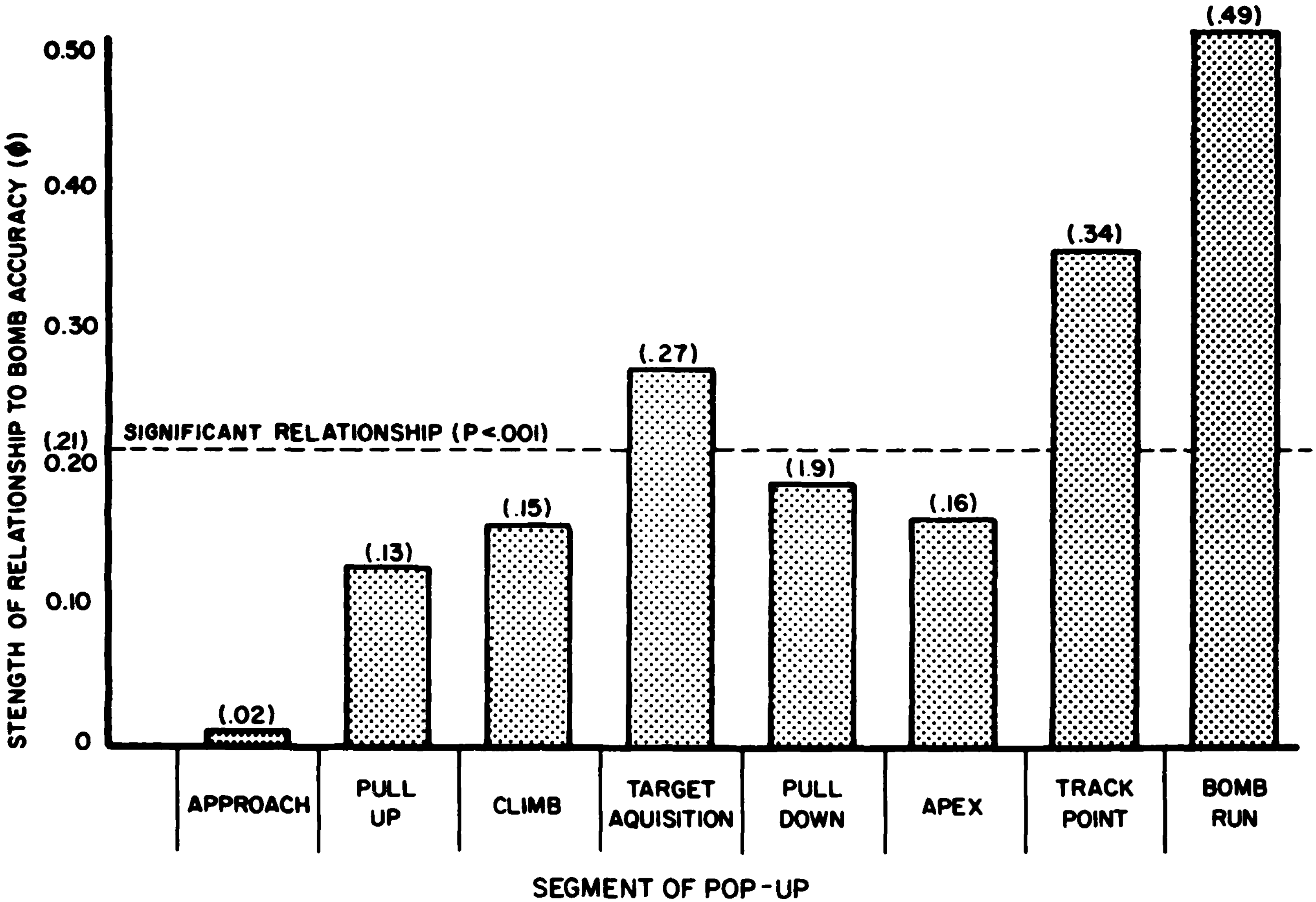

FIG. 6. LOW ANGLE (10° AND 15°) DELIVERIES.

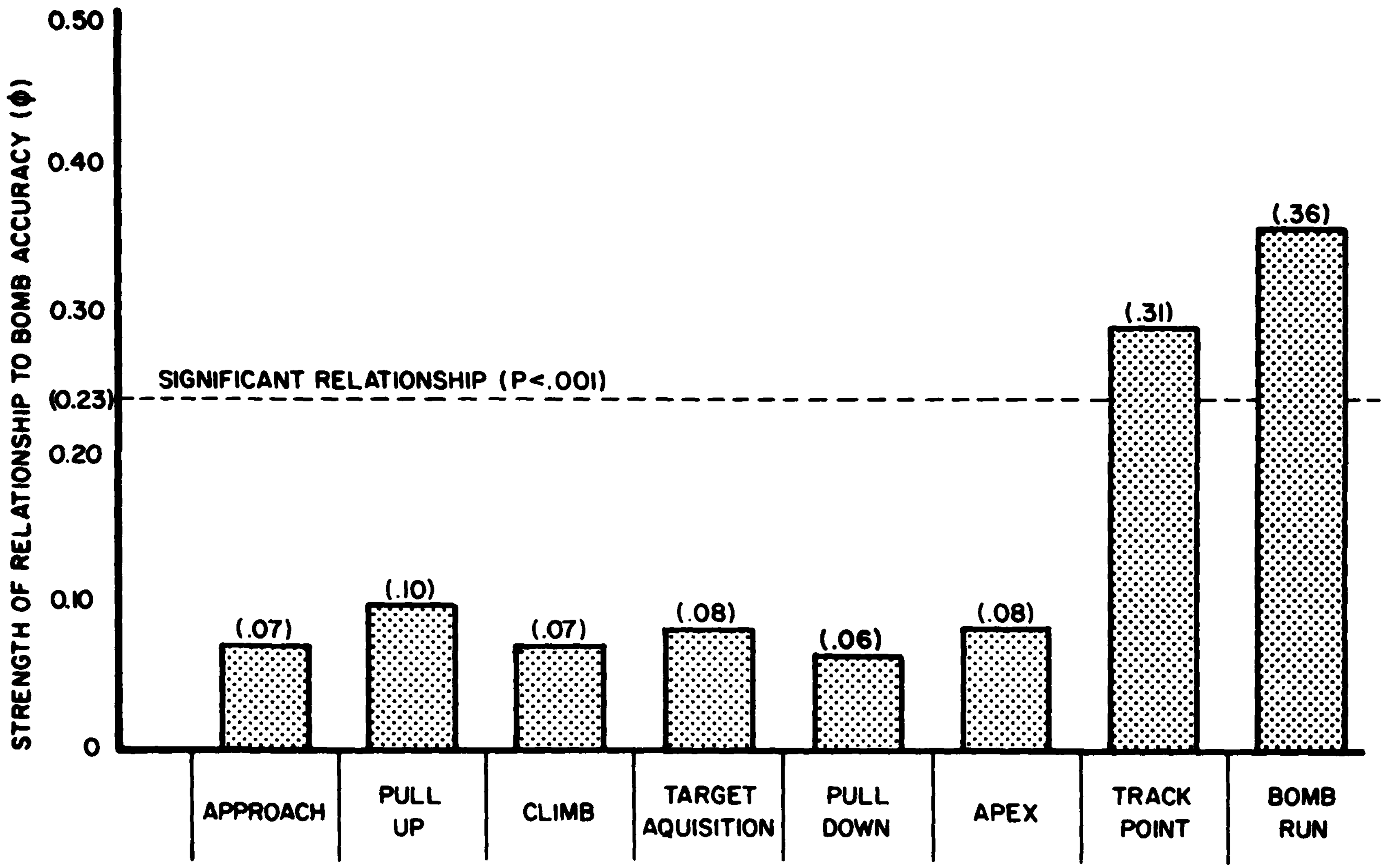

FIG. 7. HIGH ANGLE (30°) DELIVERIES.

There are several possible causes of the main results of this study. An immediate concern is that the pattern is an artifact of the ease with which different segments can be rated. It might be supposed, for example, that track point and bomb run performance ratings are more accurate and internally consistent than are the ratings of performance on other segments and that it is these rating characteristics, rather than actual segment performance, which cause the observed pattern. However, the importance of the track point and bomb run to overall accuracy appears in an analysis of reported errors as well as in the analysis of ratings. Figure 8 gives the proportion of deliveries that resulted in misses or dry or aborted passes when an error was reported for a particular part of the delivery. Since there were few errors reported for some segments, segments were pooled in ways that preserved a meaningful partition of the maneuver. The number of errors in each of the pooled segments are noted on the figure, which also shows that errors reported as occurring during the track point and bomb run are associated with an increase in the proportion of bombs off target as compared to earlier stages of the delivery. This difference is significant (^{2} = 10.0, p .0025). The severity of track point and bomb run errors can be appreciated by noting that, while the overall probability of failing to hit the target is 0.37, this probability rises to 0.80 when a specific error on track point or bomb run is reported.

These data can also give us some idea of the factors which account for less than ideal track point and bomb run execution. Figure 9 shows the relative frequency of various kinds of reported errors in track point/bomb run. This distribution of errors differs significantly from an equal-frequency distribution (^{2} = 36.8, p .001). The table shows the failure to correct properly for wind on the bomb run was the most common error, accounting for 30% of the errors reported. By contrast, wind correction was mentioned only once among the 169 errors reported for the segments preceding track point and bomb run.

The data, both ratings and errors, indicate that the track point and bomb run segments of the maneuver are the most predictive of bomb accuracy. There are at least three possible reasons for this result.

One possibility is that the greater predictivity of the final segments is due in part to earlier segments. It is possible that poor ratings on track point and bomb run reflect errors made on earlier segments as well as track point/bomb run errors. This hypothesis implies the predictivity of ratings of track point/bomb run performance should drop when only deliveries in which high ratings were given to all earlier segments are considered. Table 1 compares the predictivity of track point and bomb run ratings for all the data with that of a subset of the data which meets this condition. As this table indicates, the relationship between track point and bomb run ratings and accuracy does <u>not</u> decrease for runs in which early performance is good. If anything, the relationship decreases when earlier segments were poorly executed. Thus, early errors are not the source of the predictivity of track point and bomb run ratings.

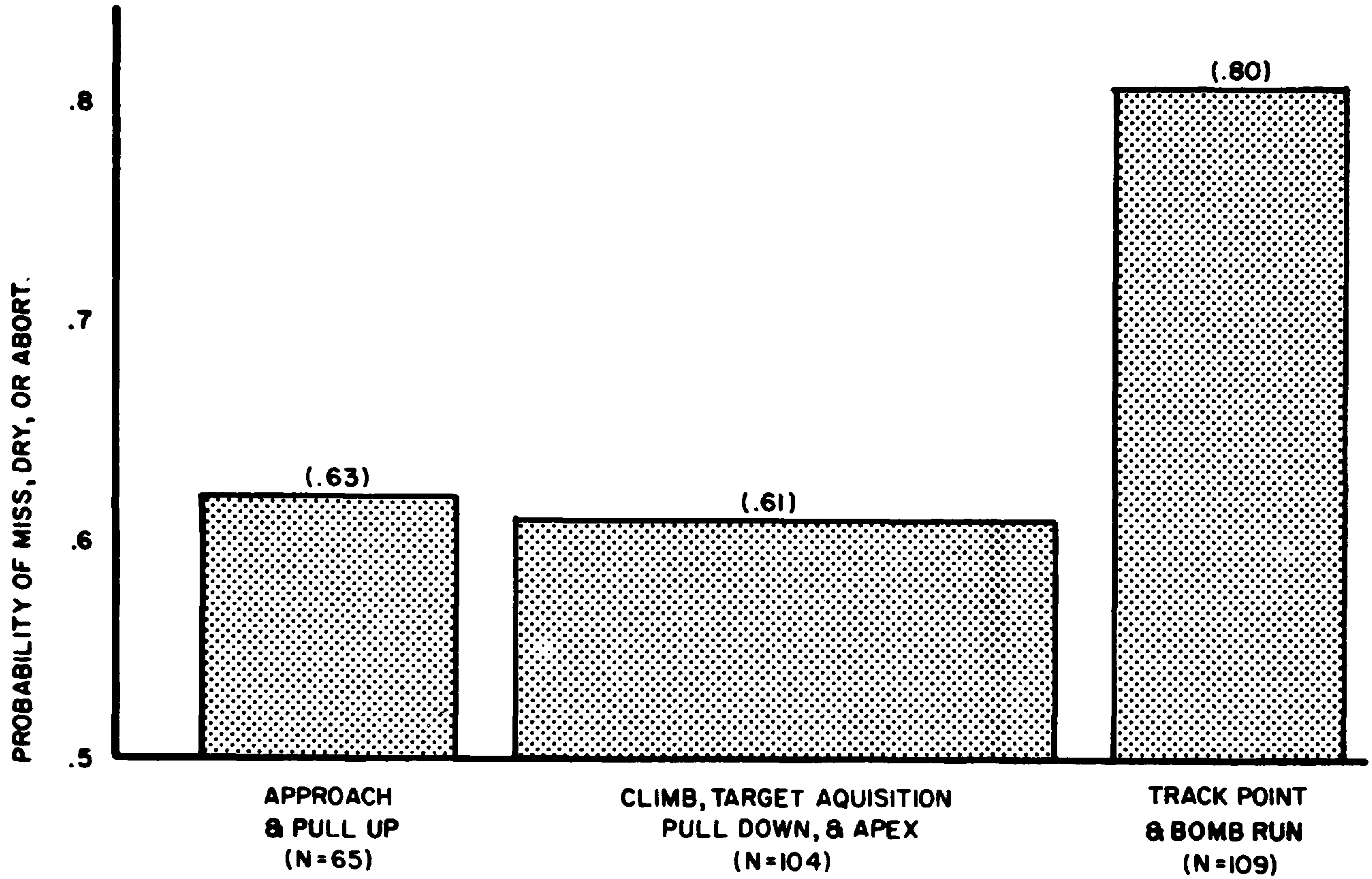

FIG. 8. RELATIONSHIP OF REPORTED ERRORS ON PARTS OF THE MANEUVER TO. DELIVERY FAILURES.

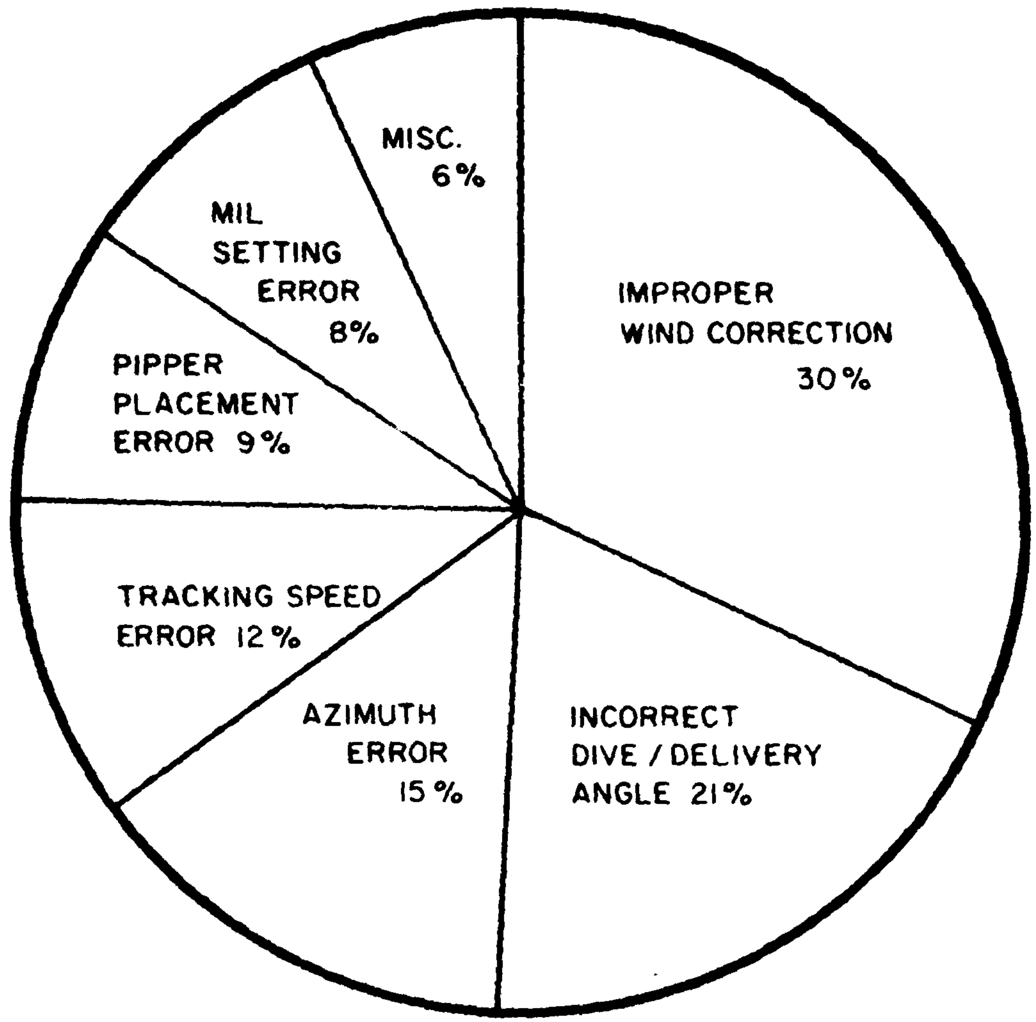

FIG. 9. RELATIVE FREQUENCY OF TRACK POINT AND BOMB RUN ERRORS.

Table 1

Strength of Relationship to Bomb Accuracy (o) of Track Point and Bomb Run Ratings (Segments 7 & 8) for Different Subsets of Deliveries

	Track Point	Bomb Run
All Deliveries	0.30	0.40
Deliveries in which all earlier segments were rated satisfactory or excellent.	0.34	0.44
Deliveries in which at least one earlier segment was rated marginal or unsatisfactory.	0.19	0.30

Two other possible reasons for the greater predictivity of these final segments are (a) they might be the most difficult parts of the maneuver to perform, and (b) independent of their relative difficulty, errors made in the final few seconds of the maneuver necessarily leave less time available for corrective action than do errors made earlier and, therefore, may be the most likely to go uncorrected and produce misses.

There is no direct index of the relative difficulty of the parts of the maneuver. Figure 10 shows the proportion of marginal and unsatisfactory ratings given for each segment. The final two segments received a significantly larger proportion of marginal and unsatisfactory ratings than did the first six segments (^{2} = 144, p .001). If this proportion is taken as an index of difficulty (Pierce et al., 1979) then the match between predictivity and difficulty is quite good. However, these ratings might be reflecting both differences in the ease with which segments are performed and differences in the pilot's criterion for acceptable performance, which in turn could be affected by the knowledge that errors in the bomb run and track point segments are difficult to correct before delivery.

Another possible index of difficulty is the number of reported errors for a given segment. Track point and bomb run account for 39% of all reported errors, significantly more than their share (^{2} = 31.5, p .001). This suggests that this part of the maneuver may in fact be the most difficult.

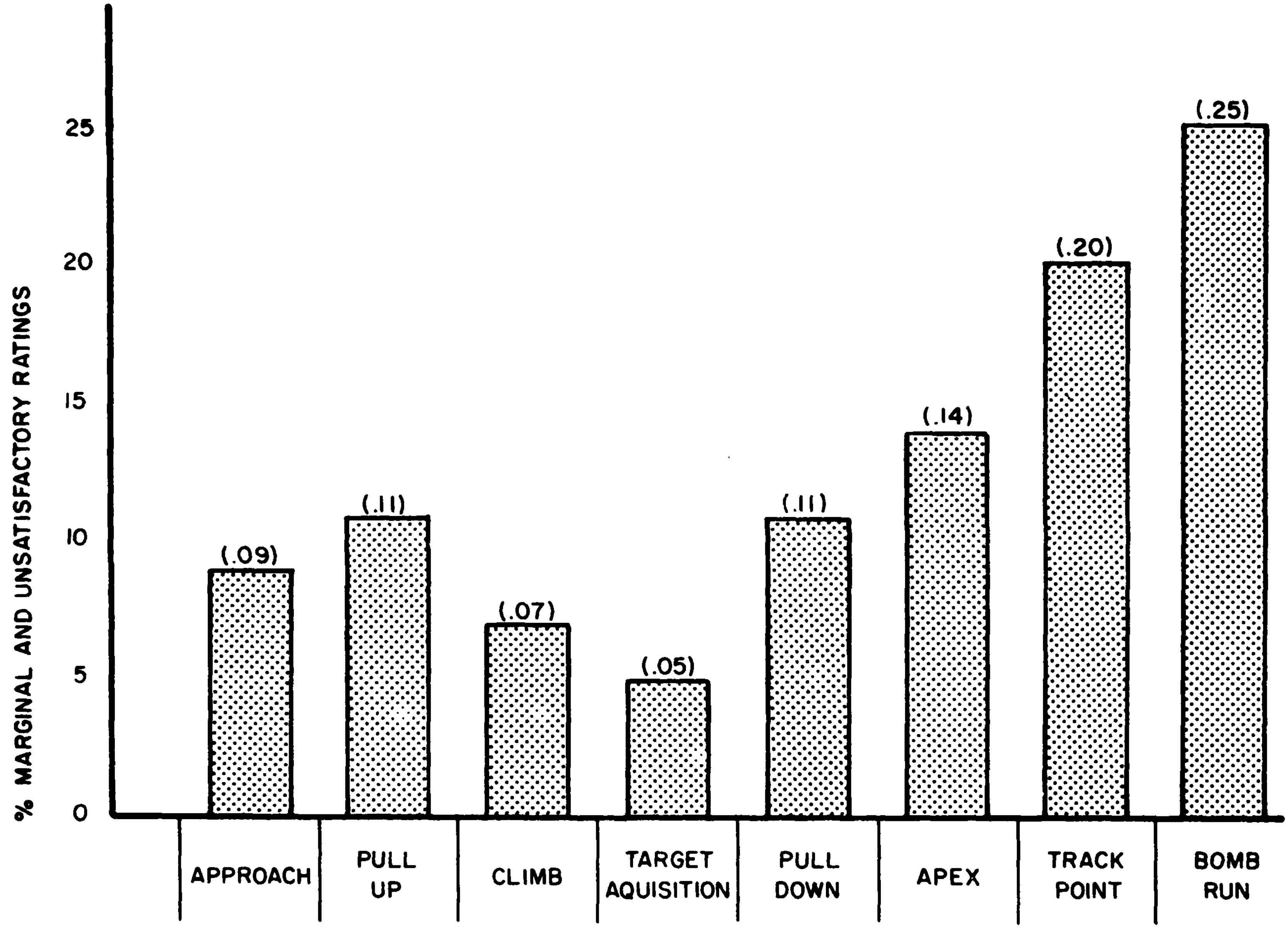

FIG. 10 PROPORTION OF MARGINAL AND UNSATISFACTORY RATINGS ON DIFFERENT SEGMENTS OF POP-UP DELIVERY.

A final piece of relevant information is the degree of independence of ratings of track point and bomb run performance. Do the contributions of these two segments to bomb accuracy reflect a common factor (such as being near the end of the maneuver) or relatively independent skills? This question can be addressed by making use of the fact that the index of association between two categorical variables, each with two levels, is equivalent to the correlation between those variables when the categories within each variable are assigned arbitrary "scores," i.e., 1 and 0 (Hays, 1973, p. 744). The particular scores chosen have no effect on the correlation, since it is impossible to produce a nonlinear transformation on two scores.

Using this equivalence, a multiple regression analysis can be performed with those variables which have already been shown to have a significant association with bomb accuracy via the chi-squared test. Table 2 presents the results of this analysis. The analysis shows that the predictive utility of track point rating is almost entirely a result of covariation with bomb run ratings. The partial correlation between track point ratings and accuracy with bomb run ratings held constant is insignificant, and the addition of track point rating to the prediction of accuracy increases the multiple correlation only negligibly. The results of this analysis are consistent with the hypothesis that both track point position and bomb run performance depend on a common factor.

Table 2

Joint and Partial Prediction of Accuracy
Using Track Point and Bomb Run Ratings

		TP	BR	A
Track Point Rating	(TP)			
Bomb Run Rating	(BR)	0.52*		
Bomb Accuracy	(A)	0.30*	0.40*	

Multiple R (TP and BR predicting A) = 0.41*

Partial R (TP predicting A holding BR constant) = 0.11*

Partial R (BR predicting A holding TP constant) = 0.29*

* Indicates significant correlation.

CONCLUSIONS

The goal of the research reported here was to determine the effectiveness of the self-assessemnt methodology explored by Project SMART. Results of a large scale test indicate that the methodology was successful in isolating the most critical elements in determining weapon delivery accuracy for the pop-up maneuver, i.e., track point position and bomb run.

The primacy of the track point/bomb run portion of the maneuver was shown both in the rating data and in reported errors (used to predict actual bomb scores). This result was highly significant and was obtained for both pilots and WSOs in each of three different F-4 wings across the country. The data also allowed the relative frequencies of occurrence of the major reported track point/bomb run errors to be categorized and compared.

It should be emphasized that these results apply only to experienced pilots on familiar training ranges, where navigation problems are minimal. It remains to be seen whether they are sustained under "high-threat," novel environments. Also, data are required from a wider range of tactical tasks, such as air combat maneuvers, to determine how skill on the pop-up maneuver fits into the overall proficiency of the mission-ready pilot.

It is clear that pilot self-assessment can be a useful source of data in identifying critical aircrew skills. The data obtained here suggest that pilots, when appropriately briefed to reduce extraneous demand characteristics of the self-assessment task, will make diligent attempts at reporting their performance accurately. And, the consistent trend across both ratings and error data suggests that pilots do reliably discriminate among individual segment performances of the pop-up maneuver.

The present results, when combined with the conclusions of an earlier examination of student weapon delivery skills using a similar methodology (Pierce et al., 1979), provide a reasonable picture of the parts of the maneuver which contribute most to overall skill at different stages of mastery. Pierce et al. concluded that, with practice, F-4 students improved their execution of the <u>initial</u> segments of the maneuver most. The final track point and bomb run segments, however, remained the most difficult parts throughout training. The data suggest that as aircrews become accurate enough to qualify as mission ready, they can execute the initial parts of the maneuver well enough so that errors in those segments do not greatly affect bomb scores. For these experienced pilots, most of the value of pop-up training drills consists of the opportunity to practice the critical final seconds of the maneuver. Therefore, it is possible that a device which simulates the rapid correction of position and aiming errors required in the final run could be of some benefit for continuation training.

The pilot self-assessment approach yielded a stable and important result in this case; however, some of the advantages that might be expected from self-evaluation techniques were not evident. For example, it was mentioned earlier that getting subjective assessments ought to be a relatively economical way to collect data since only minimal equipment is required. However, if a researcher must be present after each flight to collect the data, the expense is still considerable. During this test of the methodology, some squadrons were designated to receive full-time researcher coverage, while others were briefed on the data collection method and were asked to collect data on their own for half or all of the one month test period. The general result is that about three times the number of data forms were received when a researcher was present as compared to voluntary participation by the squadrons.

A related problem is that, even though the data forms used could be filled out in a minute or so, some pilots appeared to resent this addition to their paperwork load. This might be a problem in using the forms on a day to day basis; an even less obtrusive method would be desirable.

To conclude, the present analysis demonstrates that useful inferences about the components of task performance can be made from pilot self-assessment data; however, self-assessment may not be the most feasible method when reasonably economical and unobtrusive direct measures of performance are available.

REFERENCES

DeMaio, J. C., & Eddowes, E. E. Airborne performance measurement assessments: Low altitude tactical formation in two operating environments. AFHRL-TR-79-44. Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory, July 1980.

Hays, W. L. Statistics for the social sciences (2nd ed.). New York: Holt, Rinehart & Winston, Inc., 1973.

Nisbett, R. E. & Wilson, T. C. Telling more than we can know. Verbal reports on mental processes. Psychological Review, 1977, 84, 231-259.

Pierce, B. J., DeMaio, J. C., Eddowes, E. E., & Yates, D. Airborne performance measurement methodology application and validation: F-4 pop-up training evaluation. AFHRL-TR-79-7, AD-A072 611. Williams AFB, AZ: Flying Training Division, Air Force Human Resources Laboratory, June 1978.

APPENDIX A

MISS DISTANCE CRITERIA FOR CATEGORIZING BOMB SCORES FOR DIFFERENT EVENTS (METERS)

Category	EVENT Low Angle Bomb	Low Angle Low Drag	Dive Bomb	Dive Toss
Bull __	4.6 m	4.6 m	4.6 m	4.6 m
Hit __	32 m	53 m	44 m	50 m

☆U.S. GOVERNMENT PRINTING OFFICE: 1980- 671-143/80

AD-A091 229 DAYTON UNIV OH RESEARCH INST F/G 5/9
POP-UP WEAPON-DELIVERY MANEUVER: USE OF PILOT SELF-ASSESSMENT D--ETC(U)
OCT 80 D R LYON, J L EUBANKS, T H KILLION F33615-77-C-0054
UNCLASSIFIED AFHRL-TR-80-33 NL

2 OF 2
AD A
091 229

SUPPLEMENTARY INFORMATION

END
DATE FILMED
6-81
DTIC

SUPPLEMENTARY

INFORMATION

AD-A091229

DEPARTMENT OF THE AIR FORCE
AIR FORCE HUMAN RESOURCES LABORATORY (AFSC)
BROOKS AIR FORCE BASE, TEXAS 78235

Errata

16 JAN 1981

REPLY TO ATTN OF: TSR

SUBJECT: Removal of Export Control Statement

TO: Defense Technical Information Center
Attn: DTIC/DDA (Mrs Crumbacker)
Cameron Station
Alexandria VA 22314

1. Please remove the Export Control Statement which erroneously appears on the Notice Page of the reports listed [redacted]. This statement is intended for application to Statement B reports only.

2. Please direct any questions to AFHRL/TSR, AUTOVON 240-3877.

FOR THE COMMANDER

Wendell L Anderson

WENDELL L. ANDERSON, Lt Col, USAF
Chief, Technical Services Division

1 Atch
List of Reports

Cy to: AFHRL/TSE