NASA Technical Memorandum 81250

The Use of Total Simulator Training in Transitioning Air-Carrier Pilots: A Field Evaluation

Robert J. Randle, Jr., Trieve A. Tanner,
Joy A. Hamerman and Thomas H. Showalter

(NASA-TM-81250) THE USE OF TOTAL SIMULATOR TRAINING IN TRANSITIONING AIR-CARRIER PILOTS: A FIELD EVALUATION (NASA) 99 p HC A05/MF A01 CSCL 05I N81-16994 Unclas G3/03 41421

January 1981

National Aeronautics and
Space Administration

NASA Technical Memorandum 81250

The Use of Total Simulator Training in Transitioning Air-Carrier Pilots: A Field Evaluation

Robert J. Randle, Jr. and Trieve A. Tanner
Ames Research Center, Moffett Field, California

and

Joy A. Hamerman and Thomas H. Showalter
Ford Aerospace and Communication Corp., Stanford University, Stanford, California 94305

NASA
National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035

TABLE OF CONTENTS

PRECEDING PAGE BLANK NOT FILMED

THE USE OF TOTAL SIMULATOR TRAINING IN TRANSITIONING AIR-CARRIER PILOTS.

A FIELD EVALUATION

Robert J. Randle, Jr., Trieve A. Tanner, Jr., Joy A. Hamerman,*
and Thomas H. Showalter†

Ames Research Center

SUMMARY

A field study was conducted in which the performance of air carrier transitioning pilots who had landing training in a landing maneuver approved simulator was compared with the performance of pilots who had landing training in the aircraft. The study was accomplished at the United Airlines Flight Training Center in Denver, Colorado. NASA consulted in the study design and gathered, reduced, and analyzed the data. Forty-eight trainees transitioning to the B727 aircraft and eighty-seven trainees transitioning to the DC-10 were included in the study. It was designed and carried out in a manner that provided minimal disruption to the normal trainee flow at the center.

The study results in terms of both objectively measured performance indicants and observer and check-pilot ratings did not demonstrate a clear distinction between the two training groups. The results suggest that, for these highly skilled transitioning pilots, a separate training module in the aircraft may be of dubious value.

*Currently at Ford Aerospace and Communication Corp., Sunnyvale, California.
†Stanford University, Stanford, California.

EXECUTIVE SUMMARY

A field study was conducted in which the check-ride performance of pilots who had landing training in a flight simulator was compared with the performance of pilots who had landing training in the aircraft. The analysis of the results yielded the following findings.

Nineteen measures of performance (rating scales, scores, measured data) were acquired during the test flight. On all measures a significant overlap in performance between simulator-trained and aircraft-trained pilots was obvious from the graphed data. On the basis of these performance measures it was not possible to determine statistically which type of training the pilots had experienced. On most measures there was a small difference in favor of the aircraft-trained group. An overall statistical test of the results indicated that the differences were "nonsignificant," that is, not statistically reliable.

Safety pilots were asked to act as FAA check pilots and to provide a "pass" or "fail" for each participant in the test flight. The results were strongly influenced by the criterion utilized. When only a stringent NASA-imposed criterion, which allowed no repeats, was used (Phase I), a statistically significant failure rate (simulator-trained pilots failing more often than aircraft-trained pilots) occurred for the DC-10 trainees. When a criterion more representative of an FAA check ride, which allowed repeats of questionable performances, was added (Phase II), all trainees were deemed to have passed.

Flare and landing ratings assigned to each trainee by a NASA observer indicated that both training groups improved over the three landings of the check ride.

In Phase II a unique set of data was acquired for a comparison of simulator and check-ride performance on all trainees. The in-flight recorded data (sink rate at touchdown, vertical acceleration at touchdown, etc.) were compared with similar data recorded in the simulator. Preliminary results showed that sink rate at touchdown was slightly higher in the simulator. These data are being subjected to further analyses in a separate project, which is attempting to develop feedback techniques based on such information for use during training.

One of the parameters of interest, sink rate, was examined further in a correlational analysis. It was shown that across trainees high sink rates in the simulator were associated with high sink rates in the aircraft (and low with low). The correlation was high and statistically reliable, indicating that individual performance differences could be discerned in the simulator. A possible capability for predicting aircraft performance from simulator data using a multiple regression equation was indicated.

In a comparison of the longitudinal touchdown distance of the trainees in the NASA check-ride and revenue landings at Stapleton International Airport it

was seen that the trainees (group) dispersion was less than that for pilots on line flights. It was hypothesized that this result may have been due to a greater attention to aim-point precision in the "instructional" situation.

Trainee comments on the simulator's ability to provide landing training highlighted known weaknesses in training simulators: for example, lack of cues to depth, inability to judge sink rates, lack of peripheral cues, lack of faithful dynamic responses, and unrealistic environmental variables. These were catalogued without further analysis. All trainees rated the training as being above average. The B-727 trainees rated the training slightly lower than did the DC-10 trainees.

The landing training program for the simulator in this study was an initial effort at United Airlines and, as such, was of an experimental nature. The study results would seem to justify further effort in developing total simulation training programs.

INTRODUCTION

The use of flight simulators as substitutes for aircraft in airline pilot training has increased dramatically during the current era of the jet transport. A series of changes and exemptions to the Federal Aviation Regulations (FAR'S) to allow the increased use of simulators in training has culminated in the current regulation for Advanced Simulation (FAR 121, Appendix H), which defines the requirements for total simulation training and checking. This regulation defines three phases of simulator upgrade, each allowing progressively more critical types of training to be accomplished in the simulator, so that in the final phase all pilot training and checking may be done in the simulator.

The simulator upgrade requirements include hardware improvements to increase the fidelity of the motion and visual systems and software improvements to provide more realistic representation of aerodynamics and ground handling. Also required, although less well defined, are changes in the simulator training programs or in the ways simulators are used, including requirements for line-oriented flight training (simulation of complete missions and mission segments) and increased training requirements for simulator instructors and check airmen. These latter requirements reflect recognition of the goal of implementing the regulation: there must be complete confidence in the ability of instructors and check airmen to predict a pilot's performance in the airplane from performance in the simulator.

In spite of the previously demonstrated value of the simulator in training, complete confidence in simulator training, in the absence of an airplane check, may require that increased attention be given to the validity and reliability of pilot proficiency assessment during training and checking. Proficiency assessment will have to be objectified and standardized to increase its validity and reliability; any significant contribution that can be made in this area should increase confidence in simulator training and checking.

In anticipation of the Advanced Simulation regulation, the United Airlines Training Center and the Man-Vehicle Systems Research Division of NASA's Ames Research Center, encouraged by the Air Transport Association's Simulator Training Task Force, conducted this study of total simulator training. The study was limited to transition training (pilots moving to a new aircraft) of captains and first officers. Under the regulation for advanced simulation, transition training is permitted only after simulator upgrade according to Phase II of the regulation, although the study was conducted on simulators that would qualify only for Phase I. Therefore, the test of the simulators for training was more severe than would be allowed under the regulation. However, to insure safety in the study and on the line after the study, an airplaine check and (if needed) airplane training were provided after the "total simulation training."

The purpose of the study was to evaluate a transition training program that totally replaced the airplane with a state-of-the-art flight simulator. The evaluation procedure involved analysis of various objective measures and subjective ratings of pilot performance as a step toward objectifying and standardizing assessment techniques. The method of evaluation was to compare the performance in a standard check ride (FAR 121, Appendix F) of pilots trained totally in the simulator with the performance of pilots trained partially in the airplane in accordance with FAR 121, Appendix E. Performance measures used in the evaluation were: (1) check-pilot pass-fail ratings; (2) check-pilot ratings of specific check-ride segments; (3) a NASA-employed observer's rating of specific maneuvers; (4) trainee ratings of their own performance and of the training they received; and (5) automatically measured system variables. The statistical analysis of these data was designed to (1) compare the performance of the simulator-trained with that of the airplane-trained pilots; (2) identify any anomalies peculiar to the performance of the simulator-trained pilots; and (3) explore the possibility of developing a predictive equation of pilot performance that in the future might be used to support training and checking.

The authors wish to acknowledge the contributions of the following persons.

Personnel of the United Airlines Training Center, Denver, Colorado, were extremely cooperative in accommodating their scheduling and training activities to the needs of the study. In particular, Gary McCulloch, Manager of Flight Simulation Services, was instrumental in making the study work from its conception, for which he was mainly responsible, to its completion. John Morrison, a flight instructor at the Training Center, organized the simulator training course and worked with the authors in developing the test and data collection procedures.

Without the consultation of members of the United Airlines Master Executive Council of the Airline Pilots Association, and of the Federal Aviation Administration's Flight Standards Service, especially during the design and planning stages, the study could not have been conducted.

Captains (UAL, Ret.) Glen H. Dorward and Clifton L. Bloom served as the NASA observers, and their dedication to the study, cooperation with the authors, and general performance in carrying out their duties were outstanding. In addition, Donna L. Miller, Informatics, Inc., and Joseph G. Guercio, San Jose State University, contributed greatly to the data analysis and report editing. A special note of thanks is due Robert J. Miller, Ames Research Center, for the design, implementation, and maintenance of the on-board data acquisition and recording system. That system will be described by Mr. Miller in a separate report.

EXPERIMENTAL DESIGN

Facilities

The study was based at United Airlines Flight Training Center in Denver, Colorado. To enhance the generality of the results, two types of airplane were included: the Boeing 727 (B-727) and the McDonnell-Douglas DC-10. Two simulators were used in the study, one for pilots transitioning to the B-727 (No. 704) and one for pilots transitioning to the DC-10 (No. 605). Prior to the study, the aerodynamic programming for these two simulators was upgraded in order to receive approval by the Federal Aviation Administration (FAA) for simulator training of the landing maneuver. This upgrading subsequently qualified these two simulators for Phase I of the new regulation. The airplane training and check flights were conducted at Denver's Stapleton International Airport for the DC-10 and at Pueblo (Colorado) Airport for the B-727, the normal procedure for United. Except for the special provision that the study trainees receive all their simulator training on an approved upgraded simulator, all of the training center facilities used in normal training were used in the study.

Trainees

Captains and first officers arriving at the Training Center for transition training to either the DC-10 or B-727 were selected on a random basis to be part of the study or to receive normal transition training and checking according to FAR's 61 and 121 (including Appendixes E and F). Those trainees selected for the study were randomly assigned to either the total-simulator-training (experimental) group or the normal training (control) group. Occasionally, simulator availability modified the random assignment of trainees, either to the study or to the study groups. This modification of study procedure was necessary to minimize disruption of the regular flow of trainees of all types through the Training Center. Also, for a variety of reasons, including simulator and airplane availability, some pilots originally assigned to the study had to be dropped later, in which case they became normal transition trainees. These will be discussed in more detail later in the report. A total of 135 pilot trainees completed the study, 53 captains and 82 first officers. Of these, 48 (19 captains and 29 first officers) were transitioning to the B-727 and 87 (34 captains and 53 first officers) were transitioning to the DC-10 (see table 1).

Procedure

Trainees of both the experimental and control groups received normal ground school and normal simulator training in the appropriate landing-approved simulator without being informed of their group status. After passing their normal simulator check, the control-group trainees progressed, as normally, to receive Appendix E (FAR 121) training in the airplane. We will refer to Appendix-E-type training as landing training, since landing is considered to be the most critical part thereof. Trainees in the experimental group received their landing training in the landing-approved simulator. The simulator landing training course was developed by personnel of the Training Center and was designed to duplicate as closely as possible the standard landing training received by the control group in the aircraft.

Trainees next proceeded to the NASA check ride. For many in the experimental group, the NASA check ride was their first experience at the controls of the airplane type to which they were transitioning (B-727 or DC-10). The

NASA check ride was designed to simulate the normal check ride that would result in certification of the trainee to fly the new airplane type in revenue flights. A United Airlines check pilot served in his normal capacity in checking the first officer trainees and in simulating the role of an FAA inspector in checking the captain trainees. Check pilots were assigned to the NASA check ride on the basis of availability. The check ride consisted of the maneuvers specified in FAR 121, Appendix F, plus one additional normal landing in the following sequence: (1) taxi; (2) normal takeoff; (3) VFR approach without instrument guidance; (4) normal full-stop landing; (5) normal takeoff; (6) hooded approach, one engine inoperative; (7) missed approach; (8) VFR approach, one engine inoperative, instrument guidance available; (9) engine-out landing, touch-and-go; (10) VFR approach without instrument guidance; and (11) normal landing. The second normal VFR landing was added to provide additional data.

Upon completion of the final maneuver, the check pilot had the option of requiring or offering additional practice in the airplane before completion of the flight. This option was almost invariably exercised regardless of the trainee's performance on the check-ride maneuvers. In order to maintain his responsibility as safety pilot, the check pilot did not interrupt his monitoring of the flight to record his ratings of the trainees' performance until after the additional practice; however, it was understood that his ratings were to be based only on the check-ride maneuvers. To guard against bias in their ratings, the check pilots were not told prior to their ratings whether the trainee had received the landing training in the airplane or the simulator; that is, they were not told to which group the trainee belonged.

The original rating procedure used by the check pilots required a binary pass-fail rating, five-point ratings of the approaches and landings, and overall comments on the trainee's performance. A preliminary data analysis was conducted after 94 pilots had completed the study. Examination of their ratings and comments at that time revealed that many check pilots had interpreted the pass-fail rating as perhaps requiring a higher criterion for passing the trainee than would be expected on a standard FAA check ride. Therefore, the check pilot pass-fail rating procedure was altered for the remainder of the study. Prior to alteration, the rating asked only for a strict pass or fail with no consideration given to borderline or uncertain performance that might have occurred on some segment of the check ride. This criterion was maintained in the altered procedure, but two additional contingent criteria were added. If the check pilot issued a "fail" based on the first criterion, he was asked to judge whether the "fail" would stand if, as in a normal check ride, borderline or uncertain maneuvers could be repeated. If the trainee failed this second criterion, the check pilot was asked to consider a third: whether the trainee's performance could be considered unsatisfactory or dangerous. That portion of the study that occurred prior to the alteration in the check pilots' rating procedure will be referred to as Phase I; the portion that occurred after the alteration will be known as Phase II. Check-pilot criteria for pass-fail ratings will be discussed in more detail later in the report. Because of an increased demand for B-727 simulator usage at the Training Center, Phase II was limited to DC-10 transition training. Approximately half (41 of 87) of the DC-10 trainees participated in Phase II.

Throughout the check flight the NASA observer sat in the jump seat directly behind the captain's seat. The observer was one of two retired United Airlines captains who worked under contract with Ames Research Center. The observer's responsibility was basically to supervise data collection. In addition to scoring his own rating sheets, he installed and actuated the automatic data recording system on the airplane, and issued and collected the rating sheets of the check pilots and trainees. The observer's ratings consisted of instrument recordings and evaluative judgments made during the various maneuvers. A two-axis accelerometer was mounted on the cabin floor over the airplane's center of gravity. Vertical and lateral accelerations were recorded on an FM tape recorder starting during the approach at an altitude of 200 ft. Simultaneously, altitude was recorded from the airplane's radio altimeter. During Phase II, similar automatic recordings were also taken in the simulator.

Following the check ride the trainee completed a questionnaire about his flying history, and made ratings of both his performance in the check ride and of how well he thought his training prepared him for the check ride. Copies of the three rating forms — check pilot, NASA observer, and trainee — are presented in appendix A; explanations of the scoring procedures are also included.

After the check ride all of the collected data remained in the custody of the NASA observer until it was mailed to Ames Research Center, where it was analyzed. The data packages had no identifying trainee names; trainees were identified by numbers only.

The study was completed for the trainee when the NASA check ride was completed (about 35 min). Additional training was then given to all trainees. First officers were then certified, and captains proceeded to the FAA check ride.

RESULTS

Table 1 shows the number of trainees who participated in the study by aircraft type, training type, and position held. As was to be expected, there were certain contingencies within the scheduling process that resulted in some study participants being dropped. The B-727 scheduling was so loaded that any deviations from the planned schedule made it virtually impossible to retain the affected trainees in the study. For instance, the NASA observer had to be available on a new date, the approved simulator had to be made available for the landing training, and an instructor other than the trainee's regular instructor had to be available for a new aircraft date. Of the 68 B-727 trainees who were designated as participants, 21 were dropped for reasons given in table 2; 47 of the trainees completed the study.

Of the 21 trainees dropped from the study 9 required additional simulator training. Of the nine who required more training, only one was failed and the others were dropped for the scheduling reasons shown by the column headings in table 2. Those requiring additional training were about equally distributed between simulator and aircraft trainees.

The term "failed" here refers to failing the normal transition ground training, and the trainee could neither continue nor be in the study.

Similar information for the DC-10 is shown in table 3. Twenty-six of 113 selectees were dropped. Again, only one failure occurred. Also, both training groups were represented in the group requiring extra simulator time.

Performance Measures

Of prime interest in considering the overall results of the study was the measured and rated performance of the simulator-trained pilots versus the performance of the pilots trained in the aircraft. There were five sets of data acquired in the study: NASA-observer objective and rating-scale data; check-pilot rating-scale data; trainee ratings of the training; data measured automatically in-flight; and data recorded during the trainees' sessions on the landing-maneuver-approved simulator (DC-10, Phase II). The maneuvers flown in the NASA test flight are given in appendix A and the method of scoring the study forms of the observer, safety pilot (i.e., check pilot), and trainee are also given there. The variables measured on the aircraft were vertical and lateral touchdown acceleration and radio altitude from 200 ft altitude to touchdown. From these data the following measures were derived:

1. Descent rate

2. Descent path deviation

3. Sink rate at touchdown

4. Vertical acceleration at touchdown

5. Lateral acceleration at touchdown

6. Standard deviation of vertical acceleration at touchdown plus 2.0 sec (on runway)

7. Standard deviation of lateral acceleration at touchdown plus 2.0 sec (on runway)

For the observer data, means were calculated for the two takeoffs, three landings, and four approaches. This resulted in the eight mean measures shown in figure sets 1, 4, and 7.

Each of the four safety-pilot ratings shown in figure sets 2, 5, and 8 are mean values calculated from the six performance elements listed on his scoring form in appendix A.

The measured data are mean values based on three landings; they are shown in figure sets 3, 6, and 9.

Figure sets 1-3 show the results for the B-727 aircraft in Phase I; figure sets 4-6 are for the DC-10, Phase I; and figure sets 7-9 are for the DC-10, Phase II.

Where it was feasible to do so the scale on the abscissa of each figure was arranged to have better performance to the right. The vertical dot arrays indicate the number of trainees who received that score or rating. The group mean values are shown by the filled triangles. The number of trainees is shown on each graph. Because of some data loss due to in-flight recording equipment malfunctions, the graphs of these data show smaller trainee sample sizes than for all others. In each graph the pilots who had simulator landing training are shown at the top, the aircraft-trained group at the bottom. Also shown (circled dots) are the scores received by those trainees who were deemed to have not passed the test ride when the NASA criterion was applied by the check pilot (see the following section, Pass-Fail Ratings).

In reviewing these graphical results, a recurring finding is that the mean performance ratings for the aircraft-trained group are slightly higher than for the simulator-trained group. (Curiously, this trend is reversed in consistently higher safety-pilot ratings for the simulator-trained pilots in Phase II. See figures 8(a)-8(d). There is an obvious overlapping of scores in the two groups. Many simulator-trained pilots were rated or scored higher than aircraft-trained pilots. The overlapping indicates that the two groups are not distinct. Another way of stating this is that on the basis of the measures used in the study, one could not reliably ascertain whether any given trainee had aircraft or simulator training. In the next section an appropriate statistical test of this assertion is provided. The results are that the differences between the two groups are statistically nonsignificant; thus, for purposes of interpretation, they must be considered chance effects.

(Incidentally, the use of the term "significant" has to do with the results of statistical tests and nothing at all to do with practical import of any performance differences that may have been identified. The authors consider "reliable" and "repeatable" to be synonyms. All three refer only to the level of confidence with which probabilistic events are held to be "non-chance" or repetitive. The criteria for the practical significance of results are discernible only in real-world, operational requirements.)

Another trend shown in the graphs is the lack of a clear and consistent distinction between the scores of the "passes" and "fails." Very frequently a "fail" has a higher score than many of the "passes," both within and between the two training groups. There is a small tendency for the circles to cluster to the left (the poorer performance) end of the graph but, except for the check-pilot scores, this is not very distinct. One would expect, of course, more consistency between the check-pilot ratings and their own pass-fail judgment. Even here, however, several who failed were rated higher than some who passed. A statistical test of the differences between the "passes" and "fails" was not deemed valid due to the extremely small sample of "fails." The pass-fail results are discussed more fully below.

In the next section a discussion is provided of the statistical procedure employed to test the statistical significance of the consistent but apparently small differences between the two training groups displayed in the graphical results. Readers who wish to do so may skip that section and move on to the remainder of the report without discontinuity. The following section merely outlines the process of optimally combining the many measures for a single statistical test of the outcome. The differences were found not to be statistically reliable.

A large amount of data was taken during the course of the study. At this point it will be worthwhile to provide a rationale for the method of analysis selected. A large field study such as this required descriptive statistical procedures seldom utilized in more neatly contained laboratory studies where conditions are tightly controlled. In the latter, univariate statistical techniques are usually employed wherein a statistical test of the reliability of the difference between means or variances of the measures of a single criterion is desired and is sufficient for many purposes. However, when many variables have been measured the procedure of comparing mean scores, between the simulator and aircraft group, variable by variable is ill-advised and unproductive. There are at least the following reasons for rejecting that approach (either multiple t-tests or repeated analyses of variance).

1. On the basis of chance alone the expectation that one or more differences will result in a spurious statistical significance increases as the sample of tests increases.

2. The several univariate tests will yield no information about intercorrelation of the variables. For example, the four scoring categories of the safety-pilot ratings are adequately represented by only one of the four, because of the high correlation among the four.

3. The collection of univariate tests provides a fragmented, incohesive model of the study outcome. This collection of micro-outcomes is a poor basis for forecasting future events. That is because their contribution relative to each other is never discerned.

Given these difficulties with a score-by-score statistical analysis it was decided to employ multivariate procedures for the evaluation. Multivariate statistics provide a method for optimally combining the many variables into a single linear equation relating the predictor variables (the 19 measures) to an outcome variable. The outcome variable is taken to be group membership — the aircraft-trained or the simulator-trained group. The specific analysis selected from the multivariate family was discriminant analysis. The linear equation that results provides maximal group discrimination on the basis of the measured variables.

One of the constraints on multivariate techniques is sample size. The sample sizes in the present study were extremely small relative to the number of predictor variables. Two things were done to attempt to ameliorate this shortcoming:

1. Discriminant functions were developed for each of the data types separately.

2. Using the technique of principal components analysis the number of variables in each of the three data types was reduced.

The method of finding principal components (PC) is part of the general method of factor analysis in which data are reduced to a smaller number of underlying or latent entities each of which includes two or more of the original variables. In principal-components analysis linear combinations of the original data are found which display maximum variance and each principal component is uncorrelated with its predecessor. The first PC (PC_1) will account for most of the variance in the original data. The second will account for most of the remaining variance and be uncorrelated with the first, the third accounts for the remaining variance and is uncorrelated with PC_1 and PC_2, and so forth. The process continues until there are as many PC's as there were original variables. The analyst then uses only those first N PC's which account for a given percentage of the total variance. In the present case the observer's eight measures were reduced to five, the check pilot's four measures were reduced to one, and the measured aircraft data were reduced from seven to five. Variances accounted for in each of the three sets were 92%, 90%, and 90%, respectively, for the B-727 data. For the DC-10 data they were 91%, 84%, and 92% for Phase I and 91%, 95%, and 91% for Phase II.

The scores that entered the three discriminant analyses were these principal components. For example, the check-pilot's four rating scores were highly correlated so they reduced to a single PC or linear combination, as follows, for both training groups:

$$PC = -0.49VC - 0.50VP - 0.51IC - 0.50IP$$

where

VC = VFR control

VP = VFR procedures

IC = IFR control

IP = IFR procedures

The check-pilot's ratings were "plugged into" this equation and a single score for each trainee was calculated. This single score entered the discriminant analysis. For the observer data five PC's entered the discriminant analysis; also, for the measured data, five PC's entered.

The results of the discriminant analyses are shown in figures 10-12. The actual discriminant equations are available; however, in the interest of brevity they are not presented here. None of the functions was able to discriminate whether the trainees had had aircraft or simulator training. Tests of the statistical significance of their "power" showed that they were unreliable as discriminators. They are all near the 50% line drawn on the graph. This means that one could do as well flipping a coin as by using these performance measures in attempting to determine whether the trainee had had aircraft or simulator training.

Note also that the variance in the outcome accounted for by the several predictor variables is so low as to be almost nil. Also, the correlations between the predictor variables and the outcomes are very small.

(Many of the conventionally reported details of the analysis have been omitted. To have included them would have made this report much too large and inconvenient to read. However, much of the detailed information (e.g., the raw scores, the principal component equations, and the discriminant equations) will be retained in the files of the principal investigators for a reasonable period of time. They can be made available to interested investigators.)

Pass-Fail Ratings

Phase I- In Phase I a total of 94 trainees participated in the study. Table 4 shows the number and percentage of trainees deemed by the check pilots to have passed or failed for each equipment type and each training group. For the B-727 group the difference in pass-fail ratios between aircraft-trained and simulator-trained pilots was not statistically significant when evaluated by the chi-square statistic. For the DC-10 group the aircraft-simulator difference in pass-fail ratios was statistically significant at the $p < 0.05$ level. Almost the same percentages were obtained in the two equipment types for the simulator trained. However, with no failures in the DC-10 aircraft trained group, the estimated statistical reliability was greatly enhanced.

As stated in the Procedure section, comments by the check pilots during Phase I led to the realization that it had not been adequately communicated that the intent of the pass-fail criterion should be the same as that used in a standard FAA check ride. Some representative check-pilot comments from Phase I, which indicate this misunderstanding, are shown below for failed trainees. Each comment is from a different check ride and by a different check pilot. The training group of the trainee is indicated for each comment.

> I would have required additional successful landings which I did and the pilot did make successful landings. (B-727, simulator-trained)
>
> Passing in all respects — except reversing — used only 1.3 EPR on dry, snow-covered runway 10,000-ft length. Rest of flight very good. (B-727, aircraft-trained)
>
> Would have to repeat ILS due to runway alignment — good landings. (DC-10, simulator trained)

The following comments accompanying pass ratings (again representative) suggest a different interpretation of the criterion.

> Feeling for flare and TD were satisfactory, but additional landings after the check improved the technique noticeably . . . . (B-727, simulator-trained)

> I would have requested a repeat of the maneuver indicating that I was not satisfied with airspeed control. (DC-10, simulator-trained)

> Would have r[illegible]ed final landing for touchdown smoothness . . . . (B-727, aircraft-trained)

The apparent difference in interpretation of the pass-fail criterion across check pilots severely limits interpretation of the pass-fail ratios in Phase I.

Phase II- Table 5 shows the pass-fail results for 41 trainees in Phase II. Pass-fail Criterion A is the criterion used in Phase I. Criterion B is (in the authors' opinion) the more accurately defined FAA check-ride criterion allowing repeats of questionable maneuvers. A major difference between Phase I and Phase II is that in Phase II three of the aircraft trainees were deemed to have failed under Criterion A. Also, the simulator-trained group had a somewhat lower failure rate. Consequently, the pass-fail difference between the simulator-trained and aircraft-trained was now not significant for the DC-10. However, note that the failure rates are still unusually high (17% for the aircraft-trained and 26% for the simulator-trained), suggesting that Criterion A was still being interpreted as in Phase I, at least by some check pilots. Check pilot comments in Phase II did not show the degree of inconsistency seen in Phase I, possibly due to the addition of Criterion B.

Table 5 shows that no trainees in either training group failed under Criterion B. Since a positive answer for Criterion B precluded consideration of the third criterion, no trainee was evaluated as having been unsatisfactory or dangerous.

Considering both phases of the study and the two criteria for pass-fail ratings in Phase II, the pass-fail results taken alone suggest the following. The aircraft-trained pilots seem to have been trained to a higher level of proficiency than the simulator-trained pilots, based on Criterion A in both phases. Although the interpretation of Criterion A by the check pilots seems to have been inconsistent, one should not expect that there was more or less consistency in either training group, unless there is reason to suspect bias for or against training in either the simulator or the aircraft. Such possible bias will be discussed in the next section. However, even given a difference in proficiency between aircraft-trained and simulator-trained pilots, the results under Criterion B (Phase II) suggest that the simulator-trained were trained to a degree of proficiency acceptable for the pilots' certification (at least for the DC-10).

The implication for total simulator training is that simulator-trained pilots, as well as aircraft-trained pilots, were ready to progress to line familiarization under safety-pilot supervision. The results cannot be considered definitive in this respect, since the conditions of FAR-121 Appendix H (total training and checking in the simulator) were not exercised in the study; simulator-trained pilots did experience the airplane check before moving to the line. It could be that (under Criterion B) the NASA check ride merely confirmed passing of the simulator check and could be considered superfluous. Or it might be that some familiarization with the aircraft is necessary before unsupervised line assignment. The line familiarization experience is designed to satisfy such a requirement, and is retained under FAR-121 Appendix H.

Possible study biases- It should be emphasized that it was not possible, nor perhaps even desirable, to apply certain experimental controls in this field study. Perturbations such as weather, traffic, scheduling, and economics frequently operated to cause deviations from an ideal similarity of events for each and every trainee who participated in the study. Some of the possible sources of breaches in the original intent and ground rules of the study will now be discussed. It should be made clear, however, that the general conduct of the study and the determination of UAL training center personnel to accommodate to NASA suggested guidelines for the reduction of biases were extremely well executed.

The study was designed as "single-blind." What this meant, in this case, was that neither the check pilot nor the observer was to know whether the check ride trainee had been trained in the simulator or the aircraft. The purpose of this strategy was to obviate pre-judgment arising from a possible predisposition for the check pilot to favor one or the other type of training. The extent to which this was indeed adhered to is questionable. There are many behavioral cues revealed by a trainee who enters an aircraft cockpit for his first time at the controls, both explicit and implicit, that signal a novice. It is reasonable to assume that check pilots frequently, if not always, knew to which group a trainee belonged.

Given the possibility for bias to enter, it seemed appropriate to tabulate check-pilot responses over the course of the study to see if such was indicated. Tables 6 and 7 show this tabulation. Generally, there are too many check pilots, each having checked too few trainees, to make any statistical assertions. However, for the DC-10, check-pilot CC had a total of 13 trainees, 10 of whom had simulator training. Of the 10, he "failed" 8, or 80% over both phases. All the other check pilots over the two phases "failed" 18% of the simulator-trained pilots. This appears to be a trend; but it is not identifiable as (1) a bias against simulator training, (2) a higher criterion of performance excellence, (3) the relaxed context of "no jeopardy," or (4) some other influencing factor.

A further check was carried out. Starting with the twenty-ninth B-727 trainee and the thirty-first DC-10 trainee in Phase I, the NASA observer had been requested to provide an independent pass or fail rating for each trainee. Table 8 is a compilation of only those cases in which the check pilot and observer judgments were different. These are the only ones that were different from the last 21 B-727 and 16 DC-10 test flights of Phase I and from all 41 test flights in Phase II. Of the eight disagreements, five involved check-pilot CC. This is not to imply that the observer's determinations were the correct ones; he may have had an unconscious bias toward the simulator-trained pilots. The observer's fail rate for the 23 simulator trainees in Phase II was only 4%, considerably less than the 26% for all the check pilots, although much closer to the 9% for all check pilots except for CC. In fact, if CC's ratings are eliminated, the failure rates in Phase II are lower for the simulator-trained pilots, 9% (2 of 21), than for the aircraft-trained pilots, 17% (3 of 15). The point is not to discredit the ratings of check pilot CC; he is an acknowledged professional along with all the other check pilots. The point is to indicate how the results could be influenced if there were a bias on the part of only one check pilot, particularly with such a small sample. A second point is to suggest further evidence that all check pilots do not use the same criterion for pass-fail ratings.

Finally, a ubiquitous yet subtle source of bias was inherent in the "no-cost" or "non-jeopardy" cast of the test situation itself. There were no real-world consequences associated with mis-rating trainees. The consequences were a mark on a rating scale, limited to the study, and, except in the memory of the check pilot, made in complete anonymity. The cost was further reduced by the fact that all trainees were provided aircraft training after the NASA test flight was completed. It is not at all clear, however, whether this context would have produced more "passes" or more "fails."

Summary- To summarize, in the present study, as in any study lacking rigid laboratory controls and requiring subjective judgments of performance, there were several possible sources of bias. Nevertheless, the authors feel that the major conclusions that are suggested have not been compromised by these possible sources. Again, the conclusions suggested by the performance measures and pass-fail ratings are as follows. The simulator did train adequately for pilot certification. For some trainees it may not have trained quite as well as the aircraft. However, although the average proficiency of pilots trained in the aircraft may initially be higher than that of pilots trained in the simulator, the overlap is so great that it must be acknowledged that many pilots trained in the aircraft are not as proficient as many trained in the simulator. Finally the probability of a completely unsatisfactory or dangerous level of proficiency resulting from simulator training is, based on this study, no higher than from aircraft training; there is no evidence of either in this study.

It is important to remember that simulators and aircraft do not train independently of trainers and training programs. The training program used for total simulator training in this study was an initial attempt; it was not the result of a systematic, concentrated effort by a total training department to produce the best possible total simulator training program. Still the simulator-trained pilots as a group seemed to perform almost as well as the airplane-trained group. It is probable that a concentrated effort to produce a training program that is tailored to the advantages (and shortcomings) of the simulator will result in even higher proficiency levels. Finally, in total simulator training programs, at least at the outset, some simulator shortcomings may tend to produce slightly lower levels of proficiency; as a result, simulator and line familiarization checking may need to be intensified.

Trainee Progress During the Test Flight

Two of the NASA observer's ratings appeared frequently during data analysis in Phase I as important discriminators: the landing rating and the flare rating. It also appeared that the first landing for all trainees was frequently worse than the next two. The question to be asked was whether this was less true for those trainees who had been trained in the aircraft. Also it would be of interest to know if there was a significant improvement in the landing and flare ratings over the three landings and if this differed for the two groups. That is, was there more improvement for the simulator-trained pilots?

Table 9 shows the results of an analysis of variance of the observer's landing rating data comparing equipment types, training method, and landing number. As shown, there was a reliable difference between equipment types. Ratings were slightly higher for the B-727. There was also a highly reliable difference between the landings, the major difference occurring after the first one, for both groups. There was a small difference between the two training groups, but it was not a reliable difference. The conclusion, from these landing rating data, is that there were no reliable differences between the two training groups related to whatever practice was afforded by the test flight landing repetitions.

Table 10 shows the results using the flare rating as the criterion measure. Again there is a highly significant difference between landings for all trainees. There are no reliable effects due to equipment flown or training type. The LET interaction term, although not statistically reliable, is interesting. What it would indicate if it were slightly larger is the tendency that appears in the table of group mean scores. That is, the improvement in flare rating was more pronounced for the B-727 simulator trainees than for the DC-10 simulator trainees. This may be a reflection of the somewhat more difficult flare technique in the B-727 aircraft. As with the landing ratings, the flare ratings do not reliably discriminate between the training-group types.

The regular and reliable increase in ratings shown in these analyses may be due to the occurrence of further learning for all trainees. However, it could also be due to a "warm-up" effect in accomplishing three landings in temporal proximity. If it is the former, then learning continues even after an aircraft training session; if the latter, then perhaps even highly proficient pilots would show a similar progression.

Simulator versus Aircraft Data

The simulator and aircraft flare and landing data are being analyzed further in a separate effort. They constitute a unique and valuable set of data that provides a side-by-side comparison of simulator and aircraft performance. It cannot be properly treated in time to be included in this report. However, a gross comparison of the two sets of data is presented in figure 13. The time to flare to touchdown is shown on the abscissa as Δt. Sink rates at touchdown are on the ordinate ($-\dot{h}$). These are raw data; means of three landings were not used, as they were in previous analyses. For the aircraft, 168 landings are represented, for the simulator, 139.

Mean values for the two variables are shown by the horizontal and vertical lines. The three connected curves indicate the trend of sink rate with increasing time. They connect mean values for 1-sec intervals centered on the cardinal times. For example, the mean sink rate at 3.0 sec on the abscissa is based on all sink rates falling between 2.5 and 3.5 sec, and so forth. Curves are shown for the simulator data alone, the aircraft data alone, and for the two combined ("both"). At the point $\Delta t = 8$ and $\dot{h} = -2.0$ ft/sec there is a square. This is the point that represents the sink rate and flare interval for the DC-10 autoland system. Data for four landings of the DC-10 in the autoland mode were made available to the authors through the courtesy of the McDonnell-Douglas Aircraft Company. The landings were accomplished at Stapleton International Airport, Denver, Colorado.

The figure indicates a rather good correspondence between the distributions of landings in the simulator and aircraft; but both have large variances. The large spread of scores may have been due to the inexperience of the trainees in the "new" aircraft. The data might have looked different for experienced pilots, that is, a lower mean sink rate and a much tighter grouping of the data about that mean.

It is not intended that too much be made of the regular-appearing function that seems to emerge when means are plotted for each 1-sec interval. However, the incipient trends are interesting. Up to the 5th second, correspondence is perfect. From the 5th to 7th second they diverge; from the 7th to 8th they converge; from the 8th second onward there is a sharp divergence. Where most data lie, the agreement between aircraft and simulator is good, and both curves converge to a point very close to that resulting from the DC-10 autopilot flare law. This suggests the possible use of the autoland pitch program and altitude profile as a training model since time optimization (and distance down the runway) for minimum sink rate for pilot performance appears to occur near the autoland coordinates. Beyond the apparent optimum (8 sec) there is an increasing probability of a high sink rate again for the simulator, but not the aircraft. Data are too few here to permit conclusions to be drawn, but the difference is nonetheless provocative.

Figure 13 also shows that sink rates in the aircraft are smaller (1.6 ft/sec) than in the simulator and that Δt is larger (1.1 sec). Both of these values were statistically significant ($p < 0.0005$). However, it is difficult to imagine any training ramifications of consequence being introduced by these very small biases. The statistical significance is due to the large sample size available, that is, the large number of landings. In fact, the sample was sufficiently large that a difference between Δt's of

0.33 sec and a difference between sink rates of 0.53 ft/sec would have been "significant" at $p < 0.05$.

Figure 14 indicates that even in a homogeneous group of highly experienced pilots a range of excellence of performance is apparent. Those who had low sink rates in the simulator tended to have low sink rates in the aircraft, and high rates in the simulator went with high rates in the aircraft. (These data points are means for three landings for the 18 simulator-trained DC-10 pilots in Phase II for whom aircraft data existed.) This relationship is strong, as indicated by the correlation coefficient of 0.70, and, since the results could be expected to occur in 99 of 100 similar experiments ($p < 0.01$), they are reliable.

The equation that describes the functional relationship indicates that in predicting aircraft sink-rate performance from simulator data, the score must be increased in the ratio of 1.2 to 1.0 and a constant bias (y-intercept) added. (Since the sink-rate dimension is negatively valued, $-\dot{h}$, the positive constant will decrease the magnitude of the predicted value.) Thus, if a trainee consistently landed the simulator with a 5.0 ft/sec sink rate it would be predicted, using this regression equation, that he would land the aircraft at 3.29 ft/sec. The slope of the function is 1.2. If it had been 1.0, then prediction from simulator to aircraft would be effected by simply adding on a constant (+2.71 ft/sec), a rule of thumb.

The utility of this equation lies in its use of training data gathered in the simulator, during practice, to predict actual performance in flight. Ideally, of course, one could include more predictor variables to include more trainee performance attributes in a multiple regression equation, of which the above bivariate equation is just a special case.

The general form of such an equation is, of course:

$$y = a_0 + a_1x_1 + a_2x_2 + . . . + a_nx_n$$

Additional predictor variables (x's) would be selectable from instructor ratings, simulator measures, and other evaluative instruments currently used in the training program. Obviously, the many data types utilized in this study are not available in conventional training. Further study is required to select those variables (x's) and those weighting factors (a's) that provide the most valid prediction of the outcome score indicative of flight performance. The model itself would not change; the ingredients would evolve with increasing use and validation. The simple relationship illustrated in figure 14 indicates the validity of that particular variable. The multiple regression model is a powerful conceptual tool for combining several indices of performance and would support the instructional process. It does this by providing a mathematical analog to the instructor's cognitive process of integrating many elements into a unitary judgment. It also supports standardization through the application of a commonly used and explicit process.

The interpretation of outcome scores (y) is dependent on training goals: that score (or scores) which indicate the trainees' readiness to proceed, to whatever next step, as determined by training personnel. The conception of criterial scores is in the domain of operational requirements, well beyond the scope of this study in particular, but it is guiding it methodologically.

Trainee Comments on Simulator Fidelity

One of the more interesting aspects of the rating scale responses was the spontaneous comments written in by the respondents. There were many of these provided by the simulator trainees. They considered the simulator to be inadequate in the four categories shown in table 11. The number of comments (not trainees) is shown against the four categories. In the B-727 program 19 trainees (68%) offered 27 comments; for the DC-10, both phases, 20 trainees (41%) provided 32 comments.

Trainees in both equipment types had about the same number of comments about items in each category, but there appeared to be, perhaps, a bit more dissatisfaction with the DC-10 simulator dynamics. Perusal of the original comments reveals that no laudatory comments were submitted without qualification. The comments are given in appendix B.

Trainee Rating Scale Data

In the preliminary analysis it was determined that the trainee rating scale bore no relationship to outcomes of performance evaluation. That is, their expressed assessment of the adequacy of the training in preparing them for the test flight — their "attitudes" — was not indicative of their performance. However, because these response data were of interest in themselves they were analyzed separately. Questions of interest were the following:

1. Did captains rate the training differently from first officers?

2. Did total flying hours influence the ratings?

3. Did the simulator-trained participants feel different about the training from those trained in the aircraft?

4. Did DC-10 trainees respond differently from B-727 trainees?

Statistical analysis showed negative answers to the first two questions. In response to question (3), those trained in the aircraft tended to rate the training slightly higher than those trained in the simulator, but only for the B-727 aircraft. The difference did not reach statistical significance ($0.10 > p > 0.05$).

For the fourth question, a significant difference was found ($p < 0.05$) between the B-727 and DC-10 trainees. The B-727 trainees rated the training slightly lower than did the DC-10 trainees. Since the simulator and aircraft subgroups were lumped for this analysis, all the training was being evaluated.

Care should be taken in interpreting these results. This is a clear case of the need to evaluate the practical difference of the statistically significant result. On a scale of 1 to 5 — with 1 best and 5 worst — the B-727 pilots rated the training at a mean of 2.27; the DC-10 pilots rated it as 2.09. The difference is 0.10 points. That is a difference of less than 4.0% of the full rating-scale range. Thus, a practical difference may not be discernible. Also, the whole group of participants assessed the training as being above average in terms of its goals.

Simulator Landing Training

A key element in the study was the simulator landing training provided in lieu of aircraft landing training. The goal was to provide practice on the same kind of maneuvers to be flown in the NASA test flight. The simulator records provided to the NASA researchers (the simulator measured data) were recorded during the second of two simulator sessions. Except for practicing predetermined maneuvers, the training was frequently described as "ad lib," that is, lacking in a tightly structured syllabus. Table 12 shows the number of landings of each kind practiced by each trainee.

There is a slight tendency for the later trainees to be given fewer landings than their predecessors. Also, the total number of landings accomplished by each trainee varies considerably — from a low of 3 to a high of 16. It is not known why this occurred, but it may have been largely determined by trainee readiness and instructor judgment. However, why there is a clustering of trainees requiring fewer trials at the end is not thus explained.

In conducting transfer of training studies there are generally two training strategies. One is to train to a predetermined and measurable criterion performance. The other is to provide a predetermined number of trials on the task to each member of the experimental group. Since the simulator training in this study was frequently described as "ad lib," this presupposes the first paradigm. Since the criterion performance was subjectively determined, the readiness level of trainees may have departed considerably from homogeneity. This could have provided a sample not fully representative of the simulator capability and, thus, a less than fair test of its potential.

A fairly universally agreed upon observation on the use of training equipment is that its efficiency in effecting training is a strong function of how it is used. Factors such as the training syllabus, training personnel attitudes, pretraining and posttraining on the equipment, and trainee acceptance all contribute to the training outcome. Since the landing training program in this study was part of the experiment, it was not expected to be a finished product that would maximize training efficiency. Effort had been expended to upgrade the simulator's physical fidelity to the level of "landing maneuver approved," but this was the initial effort to develop a landing training program. The next step requires the conceptualization of a total simulator training context in which the simulator is considered as a training tool, rather than as a substitute for the aircraft.

Longitudinal Touchdown Performance

Prior to this study some data had been gathered at Stapleton International Airport pertaining to longitudinal touchdown distance from threshold for several air carriers and several equipment types. Of these, 59 were B-727's and 9 were DC-10's. Comparisons of the mean distances for those landings and the means for the present study are provided in table 13. The mean distance for the B-727 landings in this study is reliably shorter than for the revenue landings; the DC-10 landings do not differ. No interpretation of these outcomes is offered. However, the smaller dispersion of the distances in the test flight situation (the standard deviations are smaller) perhaps indicated a greater attention to aim point precision in the "instructional" context.

DISCUSSION

The two training groups in this study were seen to be statistically indistinguishable on the basis of the many performance indices that were utilized. This may not be at all surprising, given the very high experience level of the two groups of transitioning pilots. There is probably such a swift transfer of previously learned skills that any real differences have evaporated before they can be measured, or at least before they can be measured by the techniques used in the study. The quick adaptation of the trainees over the three landings, shown in the results, would seem to indicate that differences are transient. This is also supported by the check-pilot comments that occasioned the shift to the Phase II criteria for pass-fail.

Even so, the slight edge held by those who had received aircraft training motivates one to find ways in the simulator curriculum to eliminate or decrease that small difference. It seems reasonable to shift emphasis from the question of whether aircraft training is required to the question of how to maximize simulator training; it would seem uneconomical to use the aircraft in a separate training module just to erase the small differences shown in this study.

One way to increase the training value of the simulator is to set it in a training context in which the curriculum has been optimized and training personnel trained specifically to use the simulator as a training device rather than as a surrogate aircraft. Another way is to include a proficiency measurement scheme that is diagnostic, predictive, and is perceived as a training strategem rather than a trainee examination. Prediction in the simulator of ultimate performance in the aircraft would now seem to be a desired goal. The simulator data taken in this study seemed to indicate the feasibility of this goal. There is a rather high and narrow range of talent in these highly experienced pilots; but, even so, the high correlation between sink rates in the simulator and aircraft signaled that the small skill (or training readiness) differences were being sensed. That simple relationship can be expanded to include other measures of performance in a multiple regression equation, a prediction equation. It should be kept in mind, however, that validation of any quantitative model is required. One must know that an "x" level of performance in the simulator will reliably (within acceptable limits) result in "y" level of performance in the aircraft; otherwise prediction (and control) does not exist. Such a predictive model will require much future work; however, if successful, it will be extremely valuable over the long term. Side benefits to such an endeavor would be the objectification of performance and performance criteria; this would contribute to the standardization of performance requirements through the application of an explicit (versus an implicit) assessment process.

As attested to by the comments of many highly sophisticated trainees in this study, the problems of simulator realism persist: ". . . can't judge sink . . . "; ". . . poor depth perception . . ."; ". . . it doesn't feel like the aircraft . . . ." Why are sink rates higher in the simulator? It is known that these are recurrent findings in simulator research, but it is not known why. In this study a set of data was acquired that, to the knowledge of the authors, is unique. Those are the data acquired on the DC-10 simulator and on the DC-10 aircraft using the same pilots in both settings. A side-by-side comparison of these performance indices may provide insight into why there are differences between the two. Through the application of servo-theoretical models it may be possible to identify both system differences and differences due to human control technique.

APPENDIX A

FLIGHT-TEST MANEUVERS AND RATING FORMS

This appendix contains the NASA test-flight maneuvers and the observer, safety pilot, and trainee rating forms. An explanation of how each of the items was scored precedes each page requiring such explanation. The forms are those used in Phase II that were modified from Phase I, mainly in the interest of ease of scoring and to omit some items that turned out to be of doubtful relevance. The capital letters preceding each scoring explanation are those used in the main body of the text.

Note that a significant change occurred from Phase I to Phase II in the safety-pilot rating form: in Phase II the pass-fail item now included multiple criteria.

EXHIBIT 1

TEST RIDE MANEUVERS AND CONDITIONS

Sequence	Item	Details
1	Taxi	
2	Normal takeoff	1) After takeoff, if the aircraft is flying to Pueblo, the safety pilot is to assume control of the aircraft until the downwind leg of the first approach. 2) If the aircraft is at the test airport, the pilot trainee remains in control. 3) Aircraft is to maintain 15° flap configuration for the start of the downwind leg of the first approach.
3	VFR approach (No instrument guidance for pilot flying)	1) Turn onto final between OM and runway at an altitude of 1000" AGL. 2) Activate glide slope ar. localizer instruments of pilot-not-flying.
4	VFR landing	1) Full stop.
5	Normal takeoff	1) Return aircraft to 0° flap configuration for the downwind leg of the next approach.
6	Flight director engine inoperative approach (hooded)	1) Turn onto final beyond OM at 1500' AGL. 2) Fail engine on base leg. 3) Decision height is 100' AGL.
7	Missed approach	1) Return aircraft to 0° flap configuration for downwind leg of next approach.
8	VFR engine inoperative approach (instrument guidance available)	1) Turn onto final between OM and runway at 1000' AGL.
9	Engine inoperative landing (T & G)	1) Return aircraft to 15° flap configuration for downwind leg of next approach.
10	VFR approach (no instrument guidance for pilot flying)	1) Turn onto final between OM and runway at 1000' AGL. 2) Activate instruments of pilot-not-flying.
11	VFR landing	1) Full stop.

EXHIBIT 2

TRANSITION TRAINING

and

TOTAL SIMULATOR TRAINING

STUDY

DATE ______________________________

FLIGHT NO. ______________________________

PILOT NO. ______________________________

AIRCRAFT TYPE ______________________________

DATA TECHNICIAN ______________________________

SAFETY PILOT ______________________________

AIRPORT ______________________________

EXHIBIT 3

OBSERVER RATING

TAKEOFF

(A) The NASA Observer checked the appropriate letter in each of the eight boxes on top of the page. The trainee's score for Take-Off performance was determined by these check marks as follows:

Centerline Alignment: A score of 1 for Normal or a 0 for either Left or Right.

Power Management: A score of 2 for Good, 1 for Fair, or 0 for Poor.

Control Technique: A score of 2 for Good, 1 for Fair, or 0 for Poor.

Rotation Airspeed: A score of 1 for Normal, or a 0 for either High or Low.

Rotation Rate: A score of 1 for Normal, or a 0 for either Fast or Slow.

Rotation Control Technique: A score of 2 for Good, 1 for Fair, or 0 for Poor.

Initial Climb Airspeed: A score of 1 for Normal, or a 0 for either High or Low.

Heading: A score of 1 for Normal, or a 0 for either Left or Right.

These eight categories of Take-Off performance could be scored with a maximum of 11 points. The points acquired by each trainee were summed and divided by 11, thus providing a proportion of maximum possible points as a Take-Off Score.

(B) The boxes numbered one through ten were used as a rating scale with which the Observer could indicate this overall subjective evaluation of the take-off in general. A one was scored as 100% and ten was 10%.

		GW	
NORMAL TAKEOFF	W_x	V_2	
TAKEOFF	CENTER LINE ALIGNMENT L N R	POWER MANAGEMENT G F P	CONTROL TECHNIQUE G F P
ROTATION	AIRSPEED H N L (5) (-5)	ROTATION RATE F N S	CONTROL TECHNIQUE G F P
INITIAL CLIMB	AIRSPEED H N L (5) (-5)	HEADING L N R (10) (10)	

Circle one number which most nearly corresponds to your impression of the overall excellence of the takeoff. 1 is best (100%) and 10 is worst (10%).

1	2	3	4	5	6	7	8	9	10

Comments:

EXHIBIT 4

OBSERVER RATING

APPROACH

(C) The rating scale at the bottom of the page was used by the Observer as an overall subjective evaluation of the approach. A one was scored as 100% and a ten was scored as 10%.

(D) Glide Slope Deviation: The Trainee's deviations about glide slope were recorded by the Observer at 6 different points during the approach. At each point, a maximum deviation of 2 dots, high or low, was possible. The amount and direction of deviation were recorded and an average deviation score was determined for each approach.

(E) Airspeed Deviation: The Trainee's deviations about target airspeed were recorded by the Observer at 6 different points during approach. At each point, a maximum deviation of 5 knots low to 10 knots high was possible. The amount and direction of deviation were recorded and an average deviation score was determined for each approach.

	ALTITUDE	LOCALIZER	GLIDE SLOPE	TARGET AIRSPEED
ILS FINAL	OM	. . | . .	. . — . .	|---|---|---| -5 A 5 10
	1400	. . | . .	. . — . .	|---|---|---| -5 A 5 10
VFR FINAL	1100 (1000) (VFR)	. . | . .	. . — . .	|---|---|---| -5 A 5 10
	800 (700)	. . | . .	. . — . .	|---|---|---| -5 A 5 10
TAPE ON	500 (400)	. . | . .	. . — . .	|---|---|---| -5 A 5 10
	200 (100)	. . | . .	. . — . .	|---|---|---| -5 A 5 10

Approach Rating Scale

Circle one number corresponding to the category which best describes the attainment of target values over the whole approach and at missed approach altitude.

1 { always on; on at minimums
1 { frequently on; on at minimums

2-3 always on; off at minimums

4-5 frequently on; off at minimums

6-7 frequently off; on at minimums

8-9 always off; on at minimums

10 { frequently off; off at minimums
10 { always off; off at minimums

EXHIBIT 5

OBSERVER RATING

LANDING

Lateral position was not used.

(F) The rating scale at the bottom of the page was used by the observer as an overall subjective evaluation of the landing. A one was scored as 100% and a ten was scored as 10%.

(G) Longitudinal position: The Observer made a mark on the scale provided to indicate where the aircraft touched down. The number of feet from threshold that this mark represented was transcribed as down range distance.

(H) Flare: The Observer checked the adjective which best described the Trainee's performance of this maneuver. A score of one was given for Appropriate and a score of zero for either Excessive or Insufficient. For this measure only, a sum over three landings was used instead of a mean.

LANDING MANEUVER

FLARE

Excessive	_____
Appropriate	_____
Insufficient	_____

LATERAL POSITION	LONGITUDINAL POSITION
L N R	0 1000 2000 3000

Circle one number which most nearly describes your impression of the touchdown event.

1	2	3	4	5	6	7	8	9	10
Very Smooth	Smooth		Firm		Hard		Very Hard		Dangerous

Comments:

EXHIBIT 6

OBSERVER RATINGS

MISSED APPROACH

The Observer indicated a YES or NO answer to the first three questions on this page. Each YES answer scored 0.25 point and each NO answer scored 0. For the last question, GOOD scored 0.25 point and either FAIR or POOR scored a 0. Thus, the total possible points was 1.00 for the four categories.

ILS ENGINE INOPERATIVE MISSED APPROACH

ITEM	EVALUATION
Takeoff thrust at decision height.	Yes No
Flaps 25 command.	Yes No
Gear up command at positive rate of vertical speed.	Yes No
Execution of published missed approach procedures at V_2 or $V_2 + 10$.	Good Fair Poor

EXHIBIT 7

SAFETY-PILOT RATINGS

VFR APPROACH AND LANDING

(I) Control Usage: The Safety Pilot scored each listed aspect of performance using the five-point rating scale shown. A mean of these was computed.

(J) Procedure Knowledge: As above.

IFR APPROACH AND LANDING

(K) Control Usage: As above.

(L) Procedure Knowledge: As above.

These four mean scores were the input data for the statistical analyses.

The third page is the Phase I pass/fail rating form. The fourth page is the revised form used in Phase II.

Safety Pilot Evaluation Form

Safety Pilot ________________ Trainee No. ___________

This form is to be completed by the safety pilot after or during the pilot trainee's NASA test ride. Use the rating scales below and place your numerical rating in the appropriate column.

VFR Approach and Landing

Control Usage		Procedure Knowledge
_____	Rate of Descent	_____
_____	Centerline Orientation	_____
_____	Airspeed Control	_____
_____	Flare	_____
_____	Touchdown	_____
_____	Roll-Out	_____

Rating Scales

Control Usage		Procedure Knowledge
Smooth with no unwanted or erroneous inputs.	1. Excellent	All maneuvers initiated at proper time or situation.
Smooth with few unwanted or erroneous inputs.	2. Good	Most maneuvers initiated at proper time or situation.
Usually smooth with few erroneous inputs.	3. Fair	Many maneuvers initiated at proper time or situation.
Usually jerky with many erroneous inputs.	4. Poor	Some maneuvers initiated at proper time or situation.
Erratic.	5. Unacceptable	Few maneuvers initiated at proper time or situation.

IFR Approach and Landing

Control Usage		Procedure Knowledge
______	Glide Slope Tracking	______
______	Localizer Tracking	______
______	Airspeed Control	______
______	Flare	______
______	Touchdown	______
______	Roll-Out	______

Rating Scales

Control Usage		Procedure Knowledge
Smooth with no unwanted or erroneous inputs.	1. Excellent	All maneuvers initiated at proper time or situation.
Smooth with few unwanted or erroneous inputs.	2. Good	Most maneuvers initiated at proper time or situation.
Usually smooth with few erroneous inputs.	3. Fair	Many maneuvers initiated at proper time or situation.
Usually jerky with many erroneous inputs.	4. Poor	Some maneuvers initiated at proper time or situation.
Erratic.	5. Unacceptable	Few maneuvers initiated at proper time or situation.

III. Had you been asked as an FAA standards pilot to rate this trainee, would you have "Passed" or "Failed" this trainee? Circle one.

Passed Failed

Comments:

Had you been asked as an FAA standards pilot to rate this trainee, would you have "Passed" or "Failed" him/her? Circle one.

Passed Failed

If failed, in your opinion, could he/she have passed the FAA check given the traditional latitude to repeat unsatisfactory maneuvers? Circle one.

Yes No

If no, in your opinion, was his/her performance unsatisfactory or dangerous? Circle one.

Unsatisfactory Dangerous

Comments:

EXHIBIT 8

TRAINEE RATING SHEETS

(M) VFR Approach: The Trainee rated each of the three elements based on the five-point rating scale on the last page. A mean of these three scores was computed.

(N) IFR Approach, Engine Inoperative: As above.

(O) VFR/ILS Approach, Engine Inoperative: As above.

(P) Missed Approach: The Trainee rated these two elements and a mean was computed.

(Q) Landing Maneuver: The Trainee rated each of the six elements and a mean of these six scores was computed.

The remainder of the questions require no scoring explanation.

Trainee Post Test Ride

Questionnaire

Pilot No. ____________ Aircraft ___________

Date ________________

I. Regarding your performance on the NASA Test ride, use the attached rating scale to show how well the training program prepared you to perform the listed maneuvers. See rating scale at end of questionnaire.

A. Approach, Visual Reference Only, all engines.

_____ 1. Airspeed Control

_____ 2. No Reference Lateral Tracking

_____ 3. No Reference Vertical Tracking

Comments:

B. Approach, Instrument Reference, Engine Inoperative.

_____ 1. Airspeed Control

_____ 2. LOC Tracking

_____ 3. Glide Slope Tracking

Comments:

C. Approach, Visual Reference, Instrument Guidance Available, Engine Inoperative.

_____ 1. Airspeed Control

_____ 2. LOC Tracking

_____ 3. Glide Slope Tracking

Comments:

D. Missed Approach

_____ 1. Decision Height Performance

_____ 2. Airspeed Control during climb out

Comments:

E. Landing Maneuver

_____ 1. Flare Performance

_____ 2. Decrab Performance

_____ 3. Touchdown Performance

_____ (a) Lateral position at T.D.

_____ (b) Longitudinal position at T.D.

_____ (c) Sink rate at T.D.

Comments:

II. Using the same rating scale, how would you rate your performance on each of the three landings?

Comments:

Landing 1 _____

Landing 2 _____

Landing 3 _____

III. In general, how do you feel the test situation affected your flying performance? Check one.

_____ A. The test ride situation made me somewhat anxious, which hindered my ability to perform.

_____ B. The test ride situation challenged me so that I performed to the best of my ability.

_____ C. My emotional reaction to the test ride situation was minimal. My test ride performance is typical of my flying ability.

_____ D. Other. Please explain.

IV. How many flight hours, approximately, do you have as a pilot?

V. Have you ever flown the Test Ride Aircraft before as a pilot (i.e., captain or first officer)?

A. If "yes," when was your most recent experience?

B. Also, if "yes," how many flight hours as a pilot did you accumulate on that aircraft?

VI. What equipment type were you flying in your most recent set of line operations?

VII. On that aircraft, what position did you hold, captain or first officer?

VIII. During this training program, what type of landing maneuver training, simulator or aircraft, did you receive?

Rating Scale

1. The training program prepared me to perform the maneuver with no errors or misjudgments.

2. The training program prepared me to perform the maneuver with very few errors or misjudgments.

3. The training program prepared me to perform the maneuver with few errors or misjudgments.

4. The training program prepared me to perform the maneuver with some errors or misjudgments.

5. The training program prepared me to perform the maneuver with many errors or misjudgments.

APPENDIX B

SIMULATOR TRAINING: TRAINEE COMMENTS

This appendix presents, verbatim, the comments made by 17 B-727 and 19 DC-10 trainees concerning their simulator training. The comments are grouped by aircraft type and study phase. Multiple comments by a trainee are listed together. Trainee numbers are not included in order to preserve anonymity.

The touchdown itself cannot be simulated.

The simulator provides no feel for sink rate, hence no help for this phase of landing.

The failure to flare properly on the first landing resulted in a firm touchdown. This is a phase of landing which cannot be simulated.

* * *

The simulator does not allow for wind changes along the descent path. In other words, it is more mechanical in the simulator.

* * *

I feel that what I got out of the simulator for landing maneuvers could have been given in 1.5 hours.

* * *

Found variable wind in the airplane but not the sim., that is, head and tail wind on same approach.

Better visual reference in airplane.

Touchdown performance: Need practice in A/C.

* * *

All sim. periods were done at higher simulated gross weights and my power settings on A/C were slightly high for low gross which caused minor problems in getting slowed.

Flare to T.D. in sim was totally unrealistic compared to A/C.

I still wasn't sure of A/C tendencies in last few feet and nose lowering rate.

* * *

On all flight maneuvers the effect of tail skid extended on aircraft performance was not duplicated on simulator.

Ground effect in the simulator is significantly greater than in A/C.

Performance improved as I developed a feel for ground effect . . . Difference in runway width between width of runways simulator/aircraft made it slightly difficult to judge the visual glide-slope angle (vertical slot).

Landing lights used in the visual simulator had a specific cutoff point which affected visual perception during the flare portion of the landing maneuver. The simulator training appeared more effective with the landing lights off (only three or four landings were made in the simulator with the landing lights off).

* * *

I was never sure of relationship of simulator to A/C T.D. performance.

* * *

A side presentation in the visual simulator might help during training.

* * *

The simulator has no resemblance to the aircraft. Simulator very little or no flare is required to keep from balooning. A/C very different.

No X-wind in A/C but simulator seemed unrealistic.

Once again, as above, simulator tells nothing as to touchdown performance. Very unrealistic and worthless for training V/V A/C.

Long. Pos. at T.D. Simulator a little unrealistic.

Simulator gives a very false sense of smooth landing regardless of flare. No learning is evolved as far as sink rate from simulator at T.D.

First landing tried to land like simulator with little flare. Hard landing resulted.

* * *

(Pilot response to all categories on the trainee questionnaire regarding how well the training prepared him for the NASA test ride was: "No help" (10 times).)

* * *

Most difficult to determine height above runway in simulator — no problem in A/C, however.

* * *

While I think the simulator is an excellent training device, I also think it was experience that made the success or failure of the test ride; and it will be experience in the A/C that improves the aircraft handling performance (specifically landings). I can't help but wonder what these test rides would have shown if we had used initial training F/O's instead of transition.

* * *

Profile in simulator seemed steeper than A/C. I felt I was too high on several occasions as compared to actual profile in A/C.

I feel the training was adequate. I just need more time in A/C.

* * *

Simulator touchdown lacks accuracy in reflecting what would be a very firm landing in the A/C – hard to make a poor landing in the simulator.

Corrections to centerline are more difficult in simulator.

Simulator provides only very vague clues to sink in last 50 feet.

* * *

Simulator does not properly train for the rudder forces required to keep aircraft tracking on two-engine miss.

* * *

Excellent training to get the proper profile for VFR flying.

Best training is in the airplane to get the proper feel.

* * *

I find it most difficult to determine what part of my performance is based on previous experience and which part on the training received.

DC-10: PHASE I

It appeared that the pitch in the aircraft was lighter by far during the flare than the simulator.

Airplane easier to fly with engine out than simulator.

Simulator very good but aircraft some different in pitch.

Simulator height at T.D. good but some different.

* * *

SINK RATE AT T.D.: Fair/hard to judge height in simulator.

* * *

Real world much larger and airplane appears much higher from ground.

Airplane touches down much harder than sim.

Lights in sim. do not make runway appear as bright as in real world.

Firm touchdown compared to those made in sim.

* * *

The simulator is good but can't completely match actual. You use visual cues strictly for the first few landings. The simulator is pretty good in providing these.

* * *

Instructor stressed looking at end of runway during flare — impossible to do in simulator (focus at end of runway — only about 3,000 ft available on visual system).

* * *

The visual in the simulator doesn't have anywhere near enough lights (that is, ground lights).

I think the simulator brings one to on course both vertically and horizontally quicker and easier than the airplane.

* * *

Depth perception was somewhat different from simulator.

I could have made smoother touchdowns had I been more familiar with rotation rate versus sink speed of aircraft.

* * *

It's difficult to judge height in A/C because trainer lacks in presenting cockpit height above ground.

* * *

First landing was made in reference to S/O calls, just as in simulator, probably mechanical in some respects. After first landing a better understanding of the rate of flare needed was noted.

* * *

Light A/C 295 — 275,000 lb versus 315,000 in training — made for more rapid reaction requirements in aircraft.

* * *

Training did not give realistic airplane feel through flare and T.D.

* * *

Present landing simulator syllabus is worthless in my opinion. I believe that my performance would have been as good after the normal transition simulator. The concept of landing training in sim may be feasible but not with present curriculum.

* * *

In summary, the landing performance training (simulator) appears to have merit. It smoothes out the expected profiles. My own impression is that this program is excellent, provided it is integrated with <u>some</u> flying training.

DC-10: PHASE II

Interior of aircraft seemed different from simulator in regard to airspeed bleed-off with throttle position. A/C round out requires greater pitch change and control basic pressure than simulator.

* * *

I don't feel the sim training is very effectual in lateral tracking due to limited peripheral vision in simulator.

* * *

Teach a "ball park" N1 power setting of final approach configuration. Will save many periods of constant searching so more training can be spent on other important things.

* * *

Simulator does not give the feeling of mass — I tended to overshoot turns on visual approach.

Simulator didn't prepare me for the initial 50 ft above touchdown call — I couldn't believe I was only 50 ft — the flare check at 30 ft was mechanical as it was in the simulator.

Never saw a decrab performance.

From 30 ft down I felt like a spectator but seemed to improve with later landings.

* * *

Simulator just ain't the same.

* * *

Use of flight guidance covers a lot of sins.

The training was excellent from highly qualified personnel. May I suggest that the first three simulator rides be done mainly on autopilot and/or CWS so that you learn the flight guidance system first — I felt these periods were well spent in learning the simulator which slowed down, by a magnum amount, my absorption of the FGS.

Gentlemen: I am morally opposed to the idea of training on the line. If we take the historic position that passengers are entitled to a qualified crew for each trip — it is inconsistent to place a captain you say is unqualified (since he is being "shotgunned") and adding further "insult" by his never having seen the airplane he is about to fly.

* * *

TABLE 1.- TRAINING GROUP COMPOSITION

	B-727			DC-10		
	Sim	A/C	Total	Sim	A/C	Total
Phase I						
Captain	12	7	19	9	10	19
First officer	16	13	29	17	10	27
	28	20	48	26	20	46
Phase II						
Captain				9	6	15
First officer				14	12	26
				23	18	41

TABLE 2.- TALLY OF B-727 TRAINEES DROPPED AND TRAINEES REQUIRING ADDITIONAL TRAINING

Trainees dropped					
Missed assignment could not make up (sim ldgs)	Exceeded 3-day limit between sim and A/C (or LMA sim)	Simulator unavailable	Failed	Completed sim tng early and refused to stay	NASA observer unavailable
1 Sim 0 A/C	0 Sim 1 A/C	5 Sim 3 A/C	1 A/C	1 Sim 3 A/C	3 Sim 3 A/C

Trainees requiring additional training			
<3 Extra sim periods	>3 Extra sim periods	Dropped	Completed
5 Sim 3 A/C	0 Sim 1 A/C	5 Sim 4 A/C	0 Sim 0 A/C

TABLE 3.- TALLY OF DC-10 TRAINEES DROPPED AND TRAINEES REQUIRING ADDITIONAL TRAINING FOR PHASE I AND PHASE II

Trainees dropped					
Strike	Exceeded 3-day limit between sim and A/C	Same instructor not available for both lndg train sim sess	Failed	Scheduling error	NASA observer unavailable
6 Sim 4 A/C	4 Sim 5 A/C	3 Sim	1 A/C	1 A/C	2 Sim

Trainees requiring additional training			
<3 Extra sim periods	>3 Extra sim periods	Dropped	Completed
5 Sim 7 A/C	1 Sim 2 A/C	3 Sim 3 A/C	3 Sim 6 A/C

TABLE 4.- PHASE I STUDY RESULTS BY AIRCRAFT TYPE

		B-727		DC-10	
		No.	%	No.	%
Sim	Pass	19	68	18	69
	Fail	9	32	8	31
A/C	Pass	18	90	20	100
	Fail	2	10	0	0

Note: The aircraft-trained groups served as control groups.

TABLE 5.- PHASE II STUDY RESULTS

		DC-10		DC-10	
		Criterion A		Criterion B	
		No.	%	No.	%
Sim	Pass	17	74	23	100
	Fail	6	26	0	0
A/C	Pass	15	83	18	100
	Fail	3	17	0	0

TABLE 6.- B-727 PASS-FAIL RATINGS BY CHECK PILOT: PHASE I

Check pilot	Sim	A/C
A	P	
B	F,F,F	F
C	P,P	P,P,P
D	P	P
E	F	P
F	P	
G	P,F	
H	P	P,P,P
I	F	F
J	P,P,P	P,P
K	F	P
L	P,P,P	P,P
M	F	
N	F,P	P
O	P	
P	P	
Q	P	
R	P	
S		P,P
T	P	P
U		P

TABLE 7.- DC-10 PASS-FAIL RATINGS BY CHECK PILOT: ALL TRAINEES, PHASES I AND II

Check pilot	Phase I		Phase II	
	Sim	A/C	Sim	A/C
AA	P,P,P	P	P,P,F,F	
BB	P	P,P		
CC	P,F,F,P,F,F	P	F,F,F,F	P,P
DD	F,P,P,P	P,P		
EE	P,F			
FF	P,P	P,P,P,P		P,F
GG	P,F,P,F			P
HH		P,P		
II	P	P,P,P		
JJ	P	P		
KK	P			
LL	P	P,P		
MM		P,P		
NN			P,P	F,P,P,P
OO			P,P,P	P,P,P,P
PP			P,P,P	P,F
QQ			P,P,P	P,P,P
RR			P,P	
SS			P,P	

TABLE 8.- ALL CASES IN WHICH THE OBSERVER AND CHECK PILOT GAVE DIFFERENT PASS-FAIL RATINGS OVER 78 TRAINEES

A/C	Type of training	Check pilot	Observer
	Phase I		
DC-10	A/C	MM/P	F
	Sim	CC/F	P
	Phase II		
DC-10	A/C	NN/F	P
	Sim	CC/F	P
	Sim	CC/F	P
	Sim	AA/F	P
	Sim	CC/F	P
	Sim	CC/F	P

TABLE 9.- ANALYSIS OF VARIANCE OF LANDING RATINGS FOR ALL TRAINEES

Source	Sum of squares	df	Mean square	F	P
Equipment	0.034	1	0.084	4.134	<0.05
B-727					
DC-10					
Training	.043	1	.043	2.143	N.S.
Simulator					
Aircraft					
ET	.005	1	.005	.239	N.S.
Error	2.647	131	.020		
Landings	.145	2	.073	12.928	<0.001
LE	.003	2	.002	.300	N.S.
LT	.021	2	.010	1.876	N.S.
Let	.009	2	.004	.765	N.S.
Error	1.473	262	.006		

Means, % rating

	B-727		DC-10		
	Sim tr.	A/C tr.	Sim tr.	A/C tr.	Overall mean
Landing 1	77	83	76	78	78
Landing 2	84	84	81	82	82
Landing 3	83	86	80	81	82
	81	84	79	80	

TABLE 10.- ANALYSIS OF VARIANCE OF FLARE RATINGS FOR ALL TRAINEES

Source	Sum of squares	df	Mean square	F	P
Equipment B-727 DC-10	0.015	1	0.015	0.063	N.S.
Training	.726	1	.726	2.974	N.S.
ET	.213	1	.213	.872	N.S.
Error	31.971	131	.244		
Landings	4.441	2	2.220	17.473	<0.001
LE	.102	2	.051	.401	N.S.
LT	.426	2	.213	1.676	N.S.
Let	.563	2	.282	2.216	N.S.
Error	33.295	262	.127		

Means, % rating

	B-727		DC-10		
	Sim tr.	A/C tr.	Sim tr.	A/C tr	Overall mean
Landing 1	46	80	61	63	61
Landing 2	78	80	82	82	81
Landing 3	89	95	81	92	88
	71	85	75	79	

TABLE 11.- ARBITRARY CATEGORIZATION OF TRAINEE SPONTANEOUS COMMENTS ON SIMULATOR INADEQUACIES

Category	Number of comments	
	B-727	DC-10
Flare	12	12
Height estimation		
Depth estimation		
Sink rate		
Visual geometry		
Dynamics	3	9
Mass		
Response		
Feel		
Gross weight		
Environmental	4	4
Wind		
Turbulence		
Ground effect		
Random natural phenomena		
General	8	7
Sim inadequate		
Sim easier to fly		
A/C easier to fly		
Sim syllabus inadequate		
Sim session too long		
Philosophical reservations		

TABLE 12.- DC-10 SIMULTAOR LANDINGS BY TRAINEE AND TEST-FLIGHT MANEUVER

Trainee	VFR-1	VFR-2[a]	VFR-3	Total landings
1	12	4		16
2	8	3		11
3	7	2		9
4	11	1	1	13
5	1		4	5
6	3	1	2	6
7	3	2	1	6
8	3	2	1	6
9	2	2	2	6
10	1	1	5	7
11	1	1	2	4
12	1	1	1	3
13	1	1	1	3
14	1	1	1	3
15	2	1	1	4
16	2	1	1	4
17	i	1	1	3
18	3	2	2	7
19	4	3	1	8
20	2	1	1	4
21	2	1	1	4
22	3		1	4
23	3	1	1	5
Total	77	33	31	141

[a]One engine out, instrument guidance available.

TABLE 13.- COMPARISON OF LONGITUDINAL TOUCHDOWN DISTANCE OF REVENUE FLIGHTS AND NASA TEST FLIGHTS

	Revenue	Test flight	Difference	p
B-727				
N	59	42		
Mean	1721.66	1316.38	405.28	<0.005
Standard deviation	470.00	238.90		
DC-10				
N	9	84		
Mean	1611.11	1767.48	156.37	N.S.
Standard deviation	561.68	447.68		

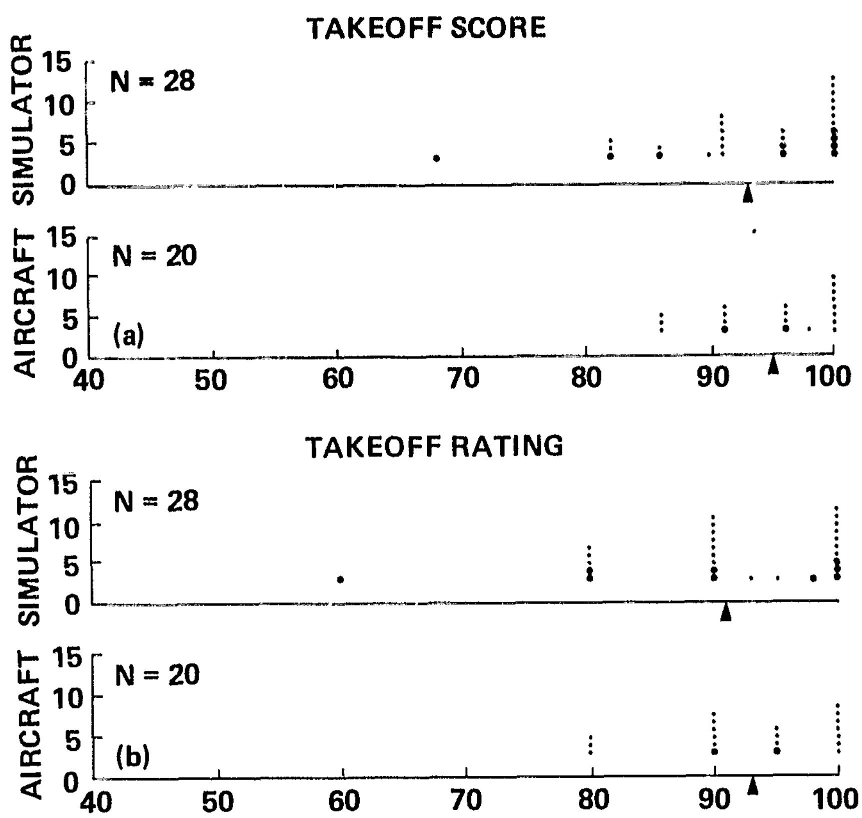

Figure 1.- Phase I observer scores for B-727 trainees.

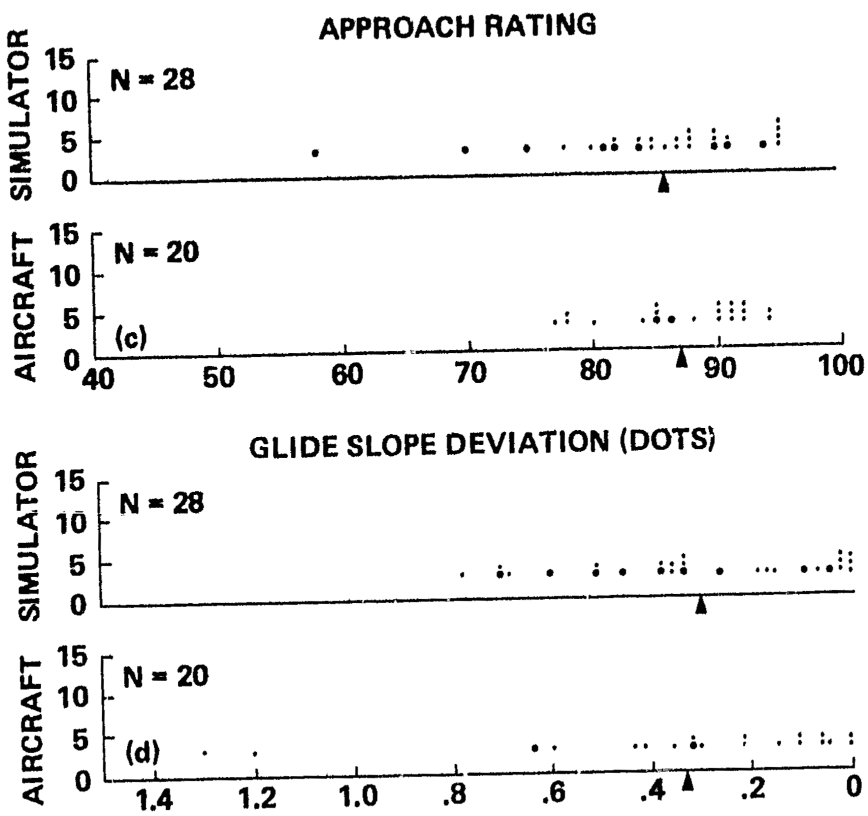

Figure 1.- Continued.

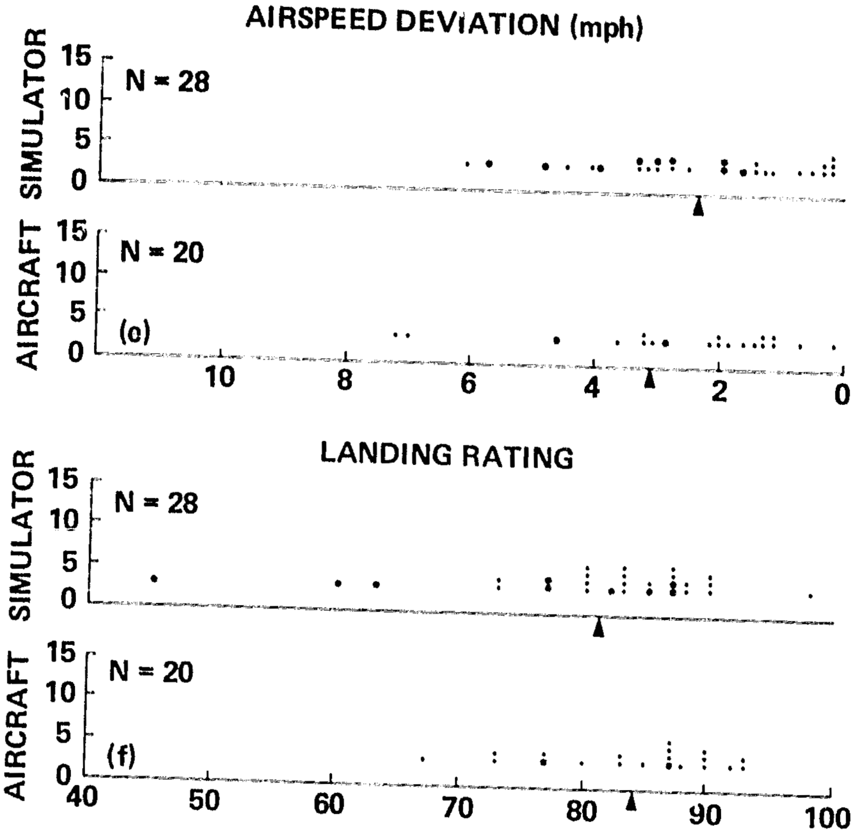

Figure 1.- Continued.

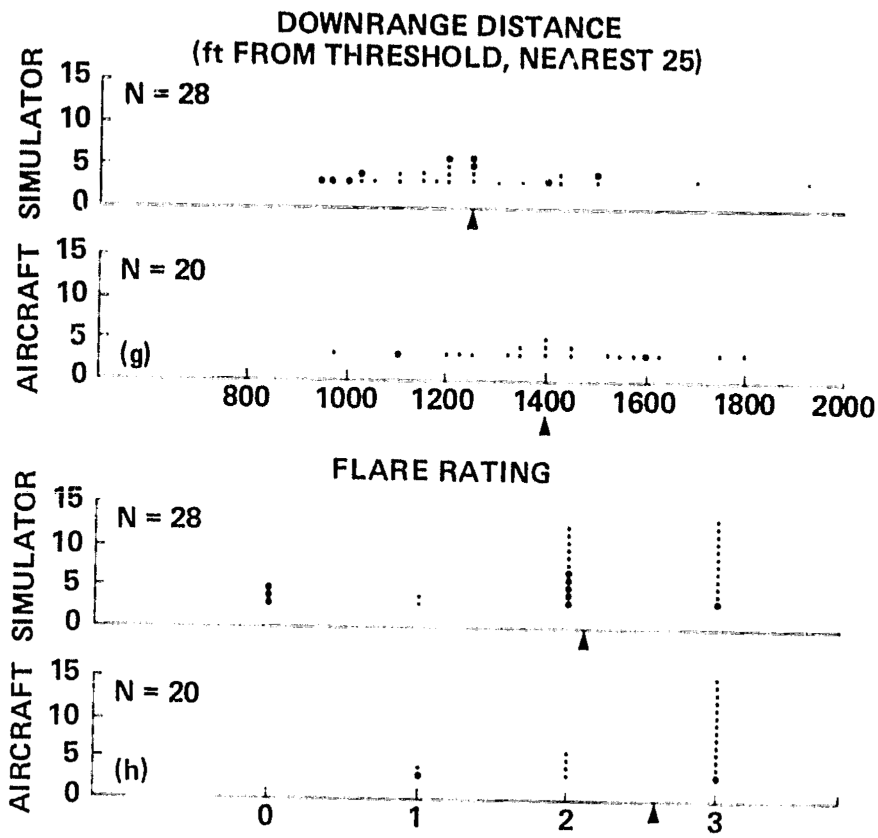

Figure 1.- Concluded.

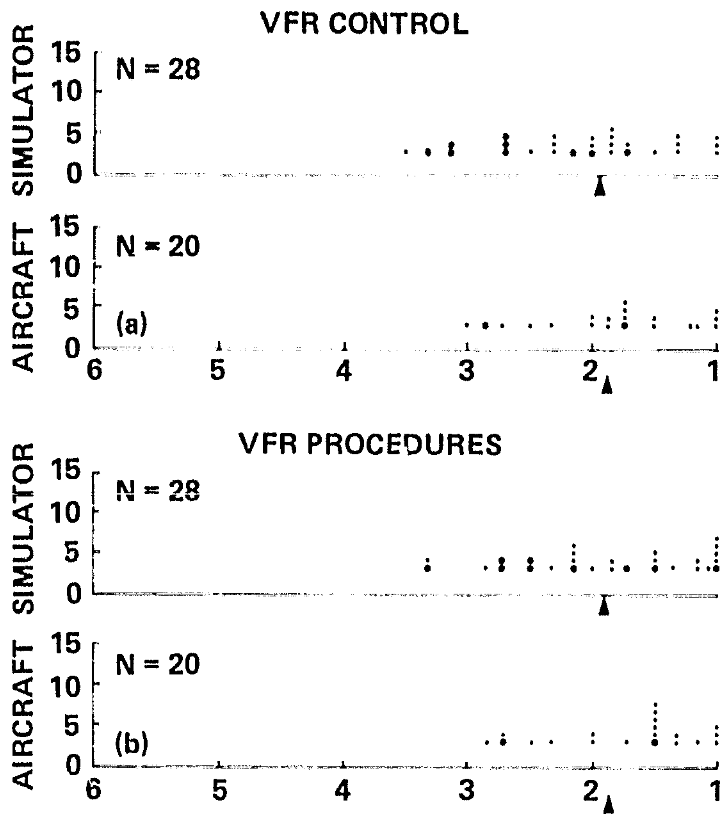

Figure 2.- Phase I pilot scores for B-727 trainees.

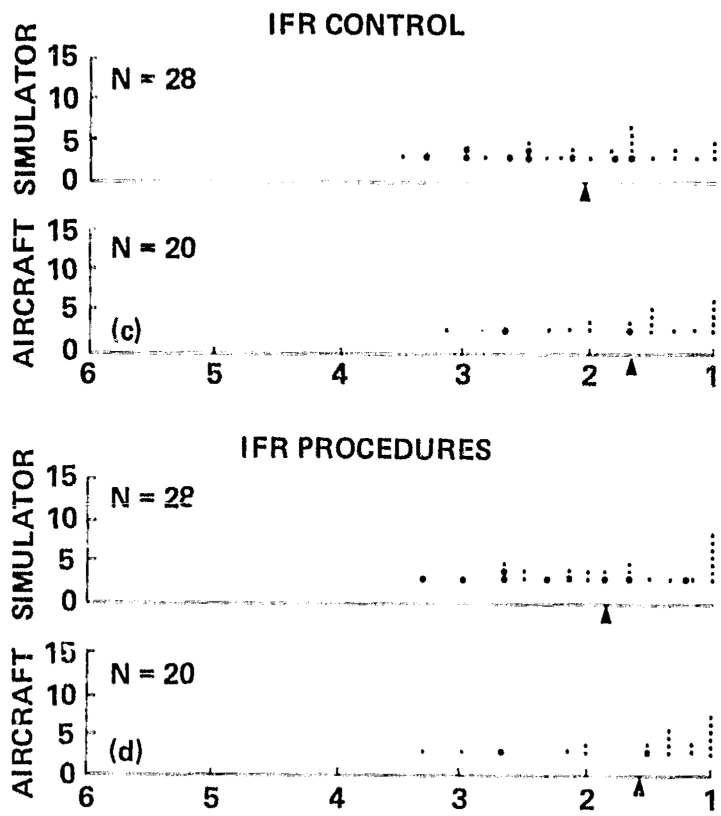

Figure 2.- Concluded.

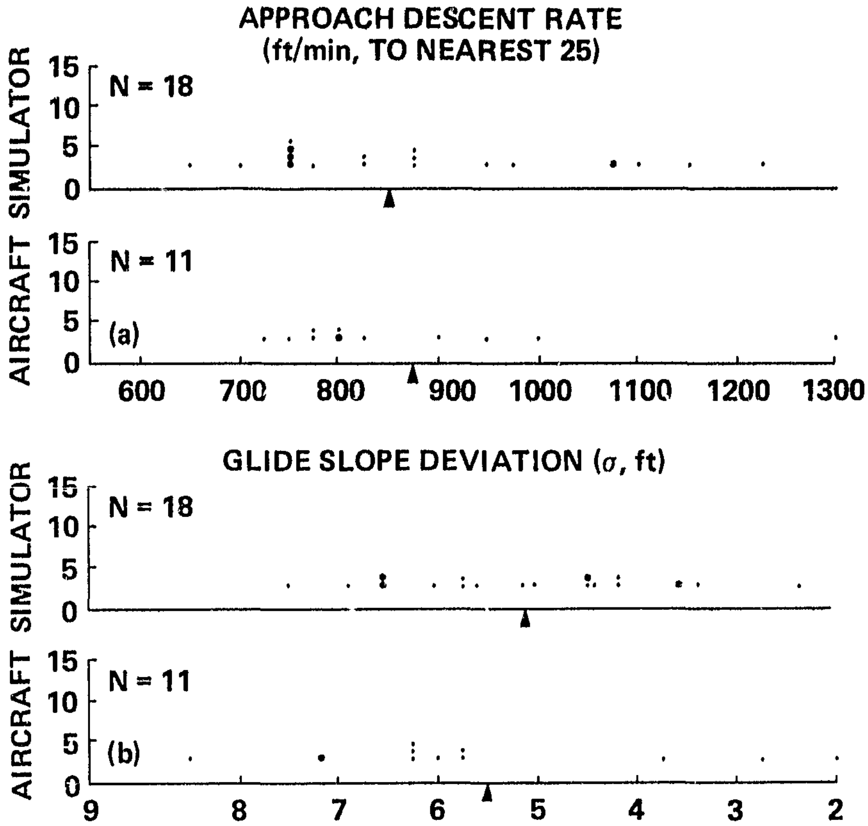

Figure 3.- Phase I aircraft-measured data for B-727 trainees.

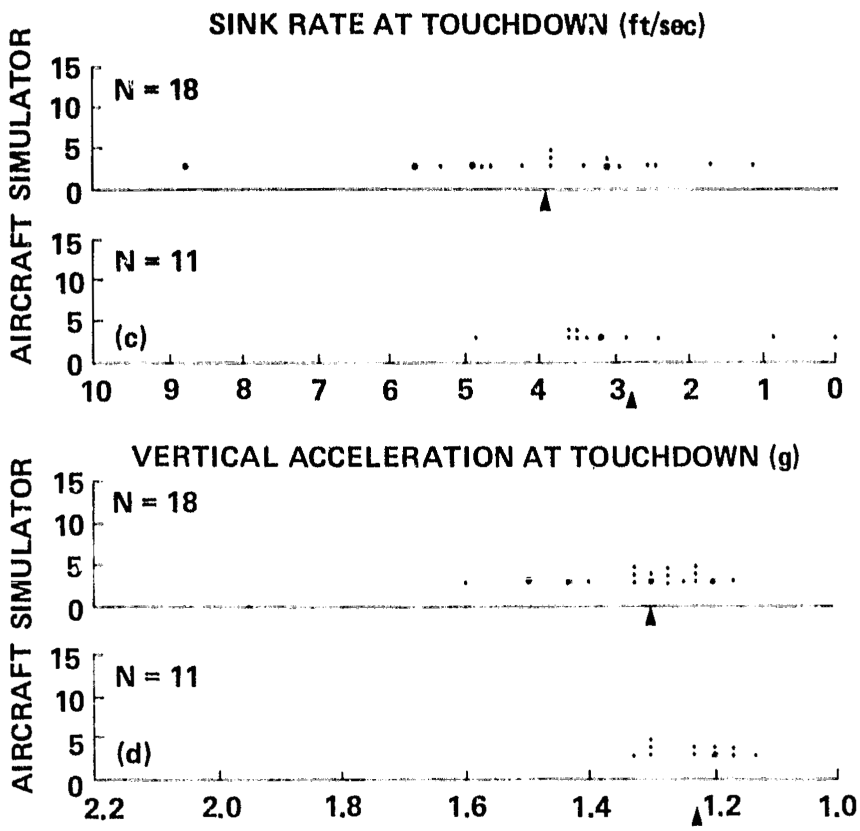

Figure 3.- Continued.

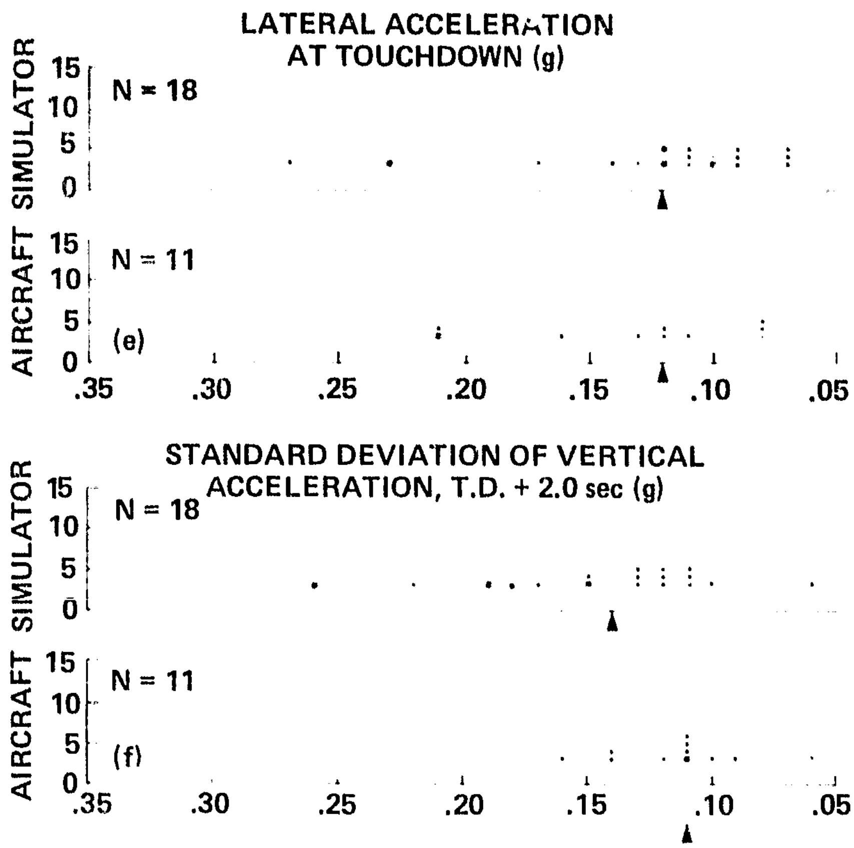

Figure 3.- Continued.

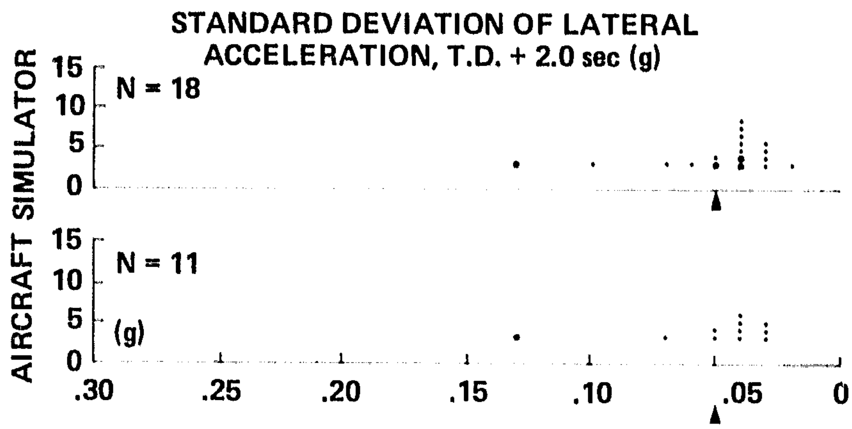

Figure 3.- Concluded.

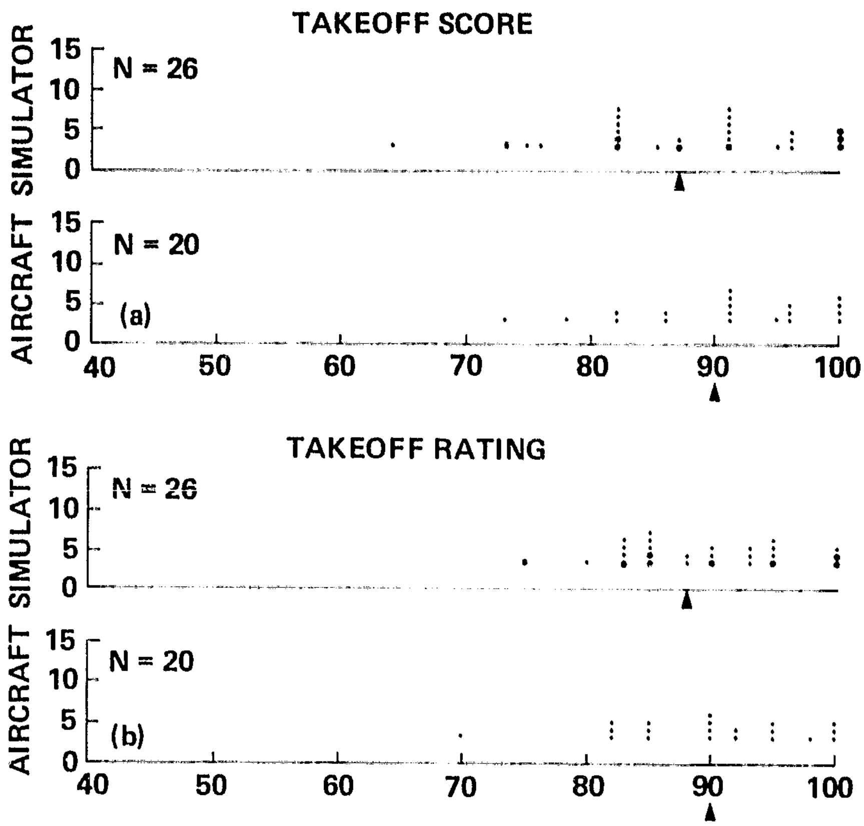

Figure 4.- Phase I observer scores for DC-10 trainees.

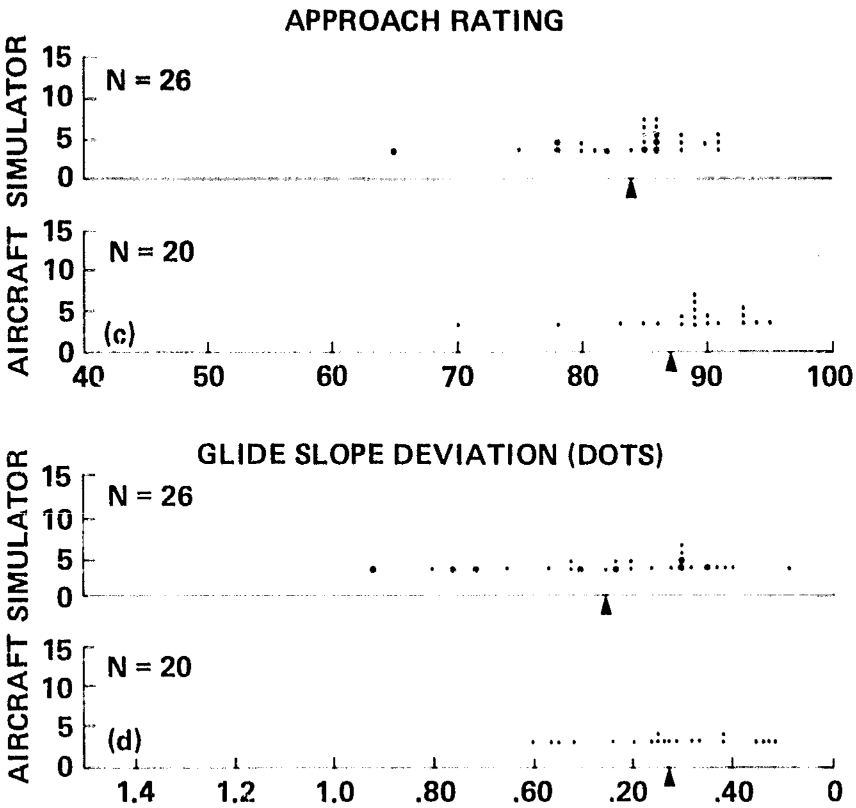

Figure 4.- Continued.

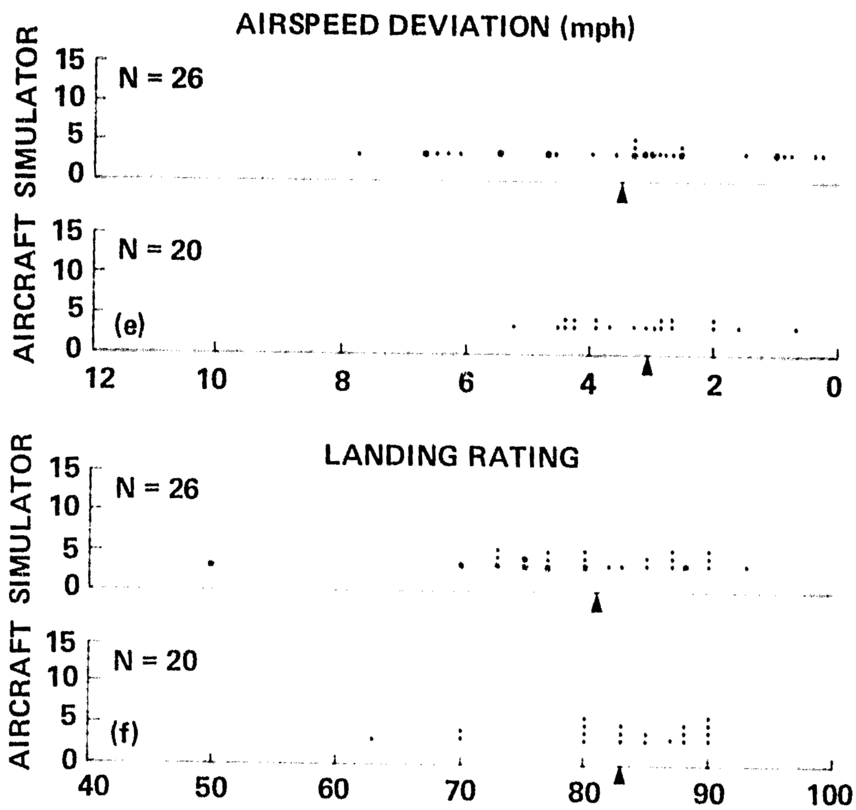

Figure 4.- Continued.

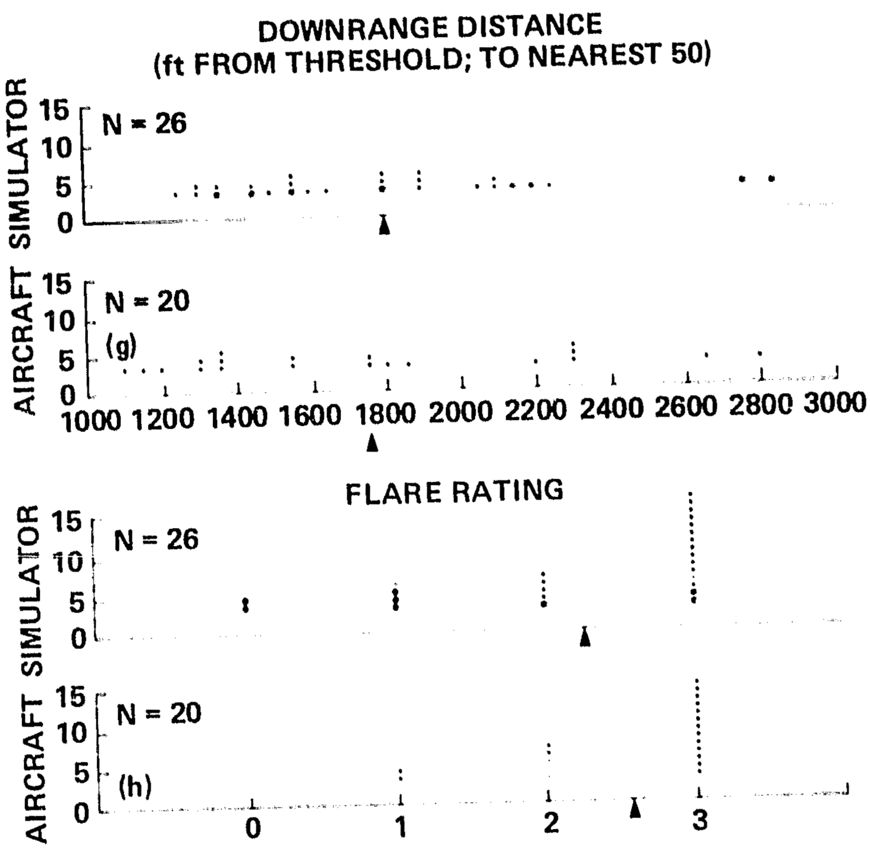

Figure 4.- Concluded.

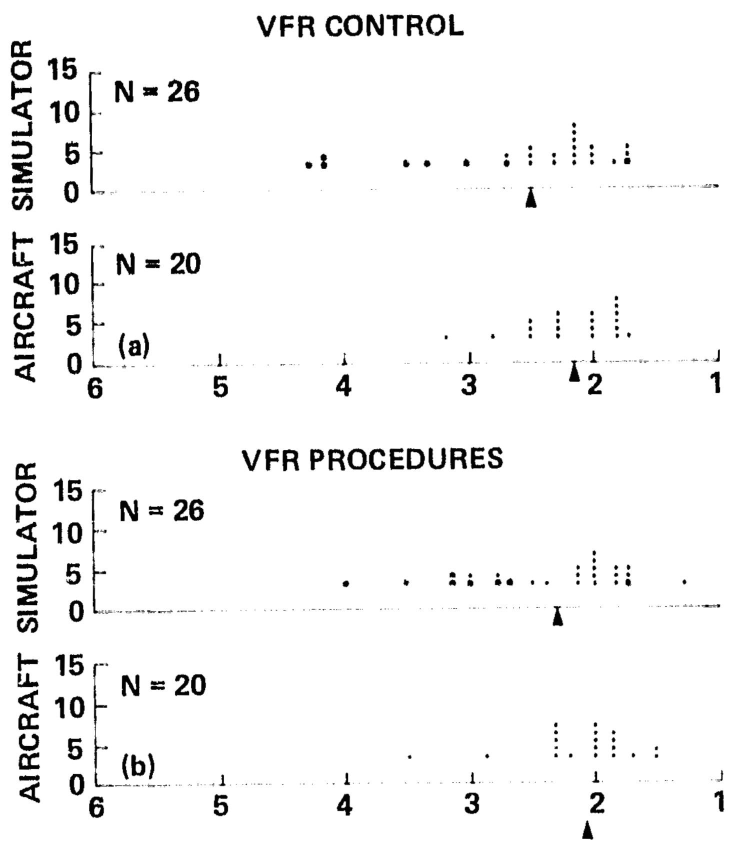

Figure 5.- Phase I check-pilot scores for DC-10 trainees.

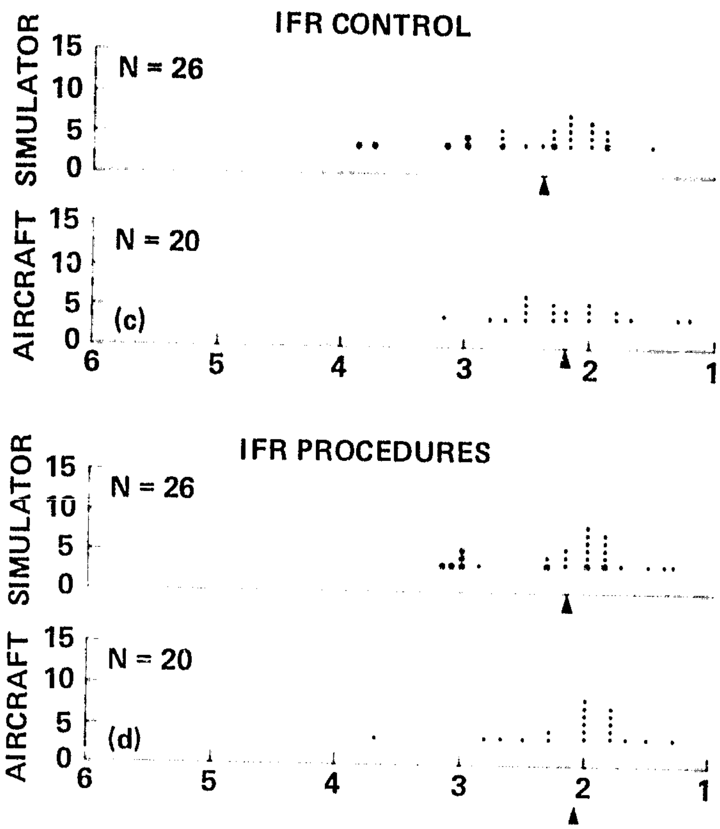

Figure 5.- Concluded.

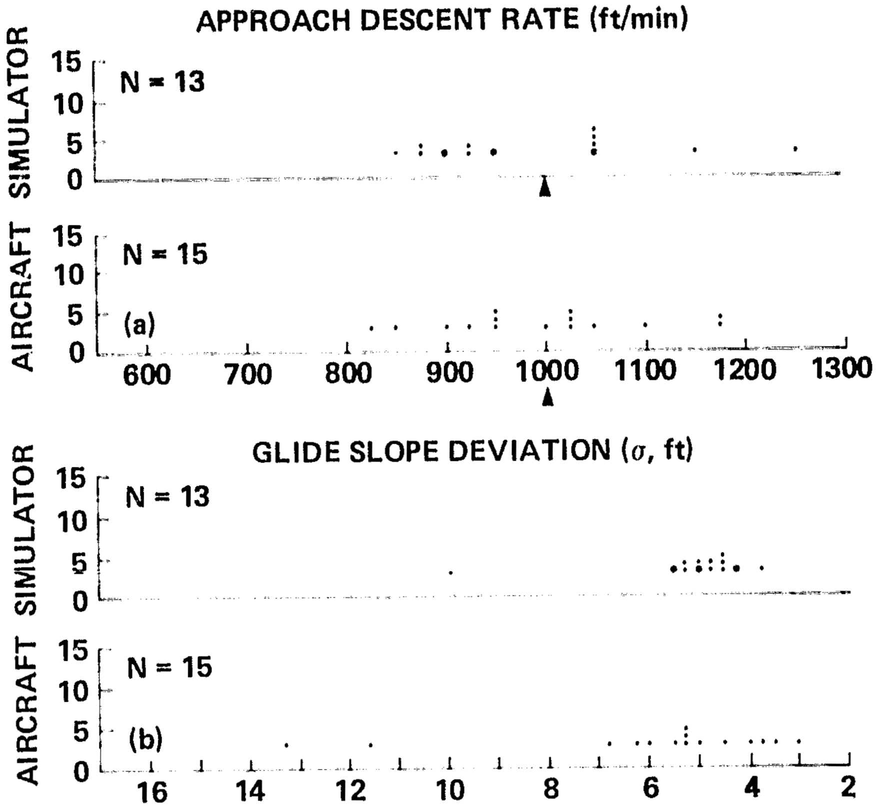

Figure 6.- Phase I aircraft-measured data for DC-10 trainees.

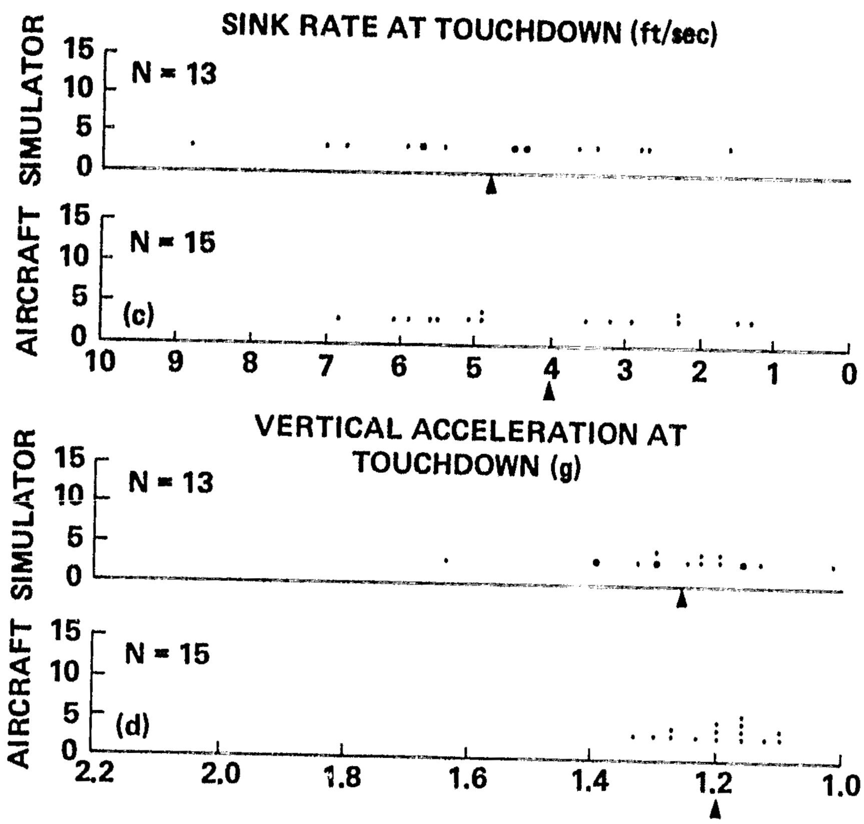

Figure 6.- Continued.

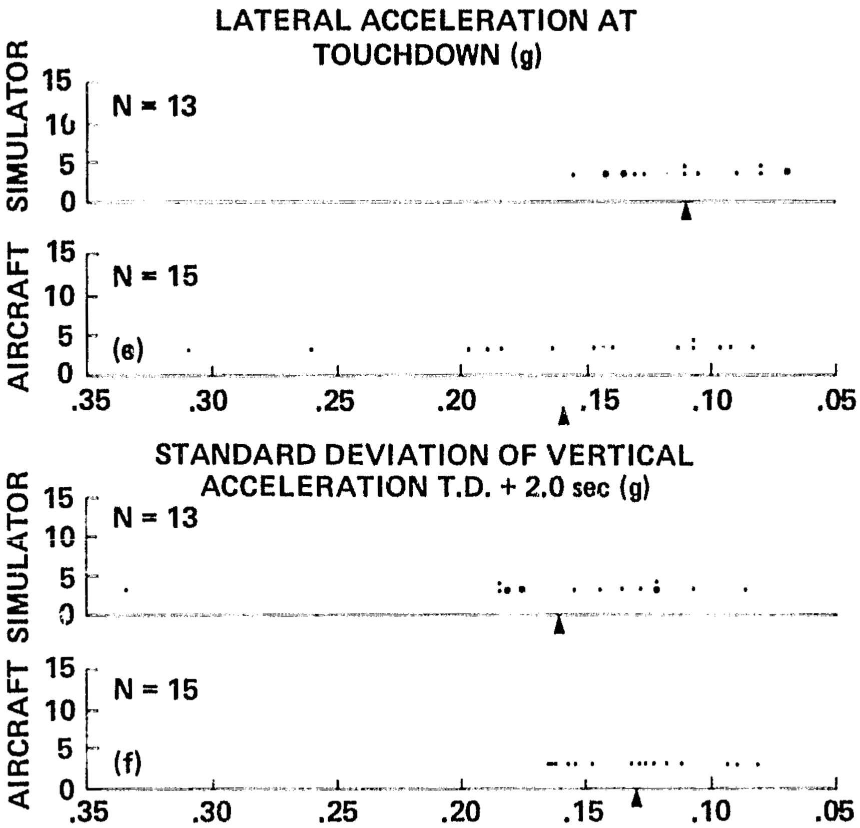

Figure 6.- Continued.

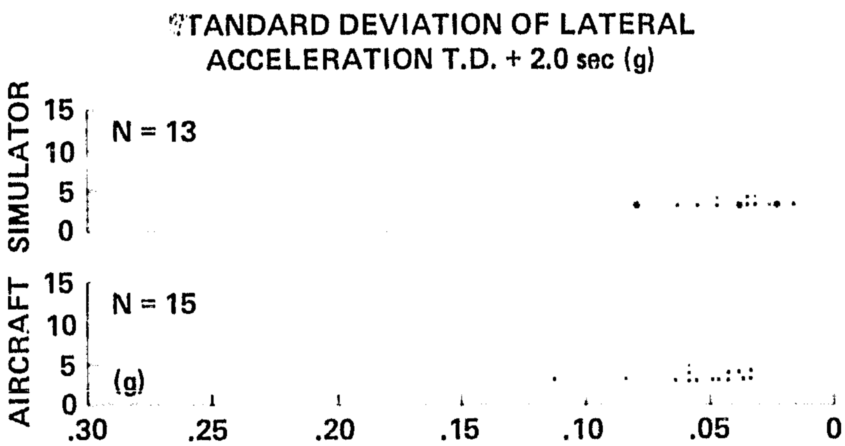

Figure 6.- Concluded.

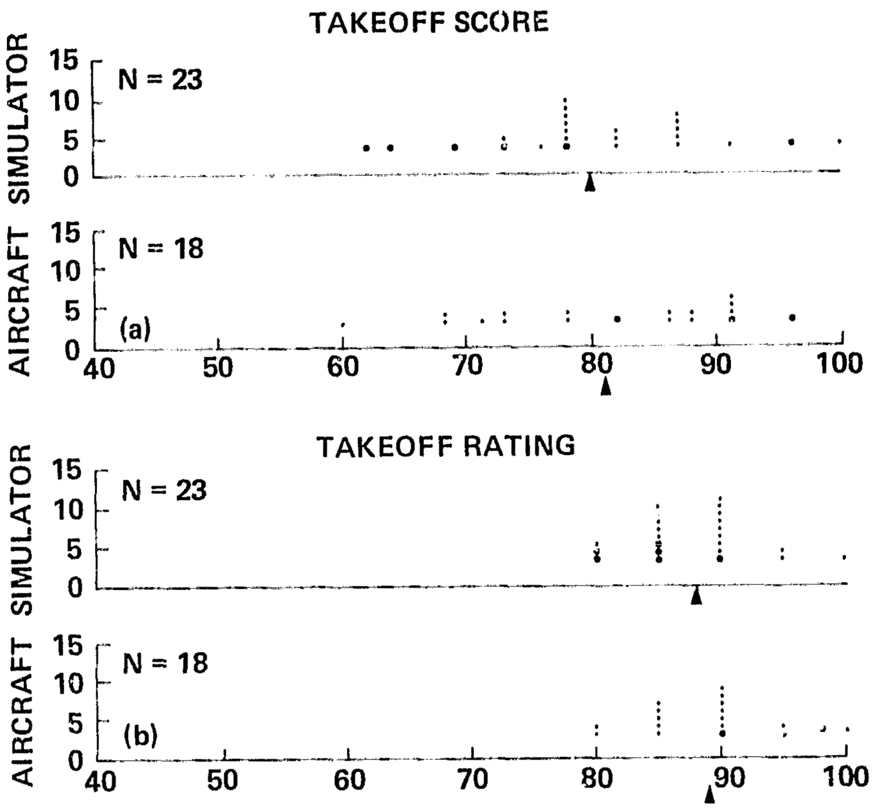

Figure 7.- Phase II observer scores for DC-10 trainees.

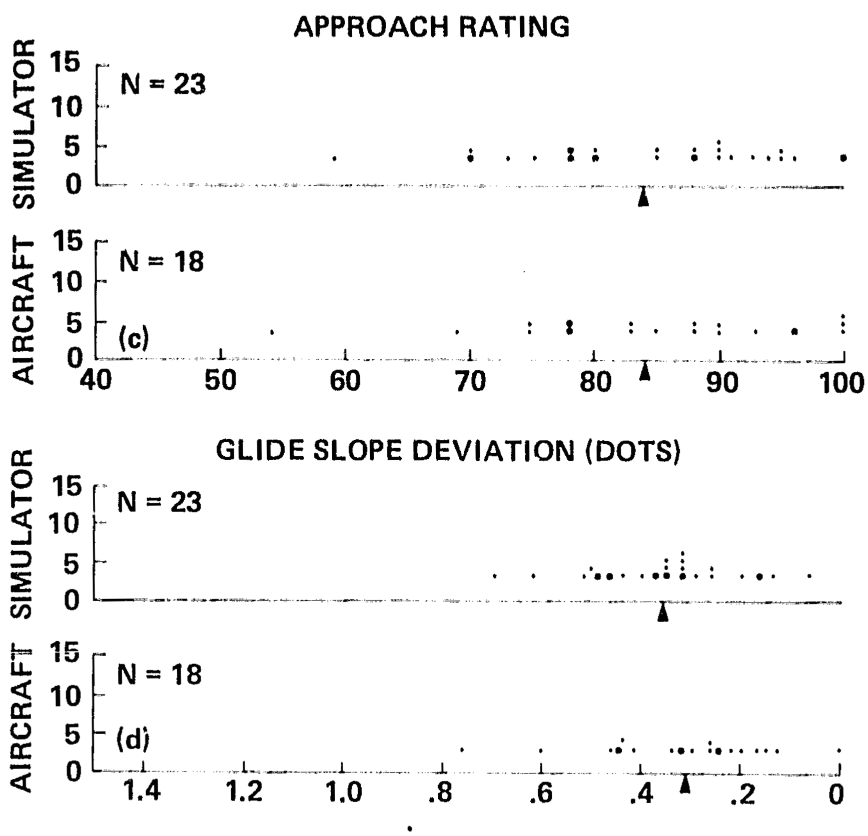

Figure 7.- Continued.

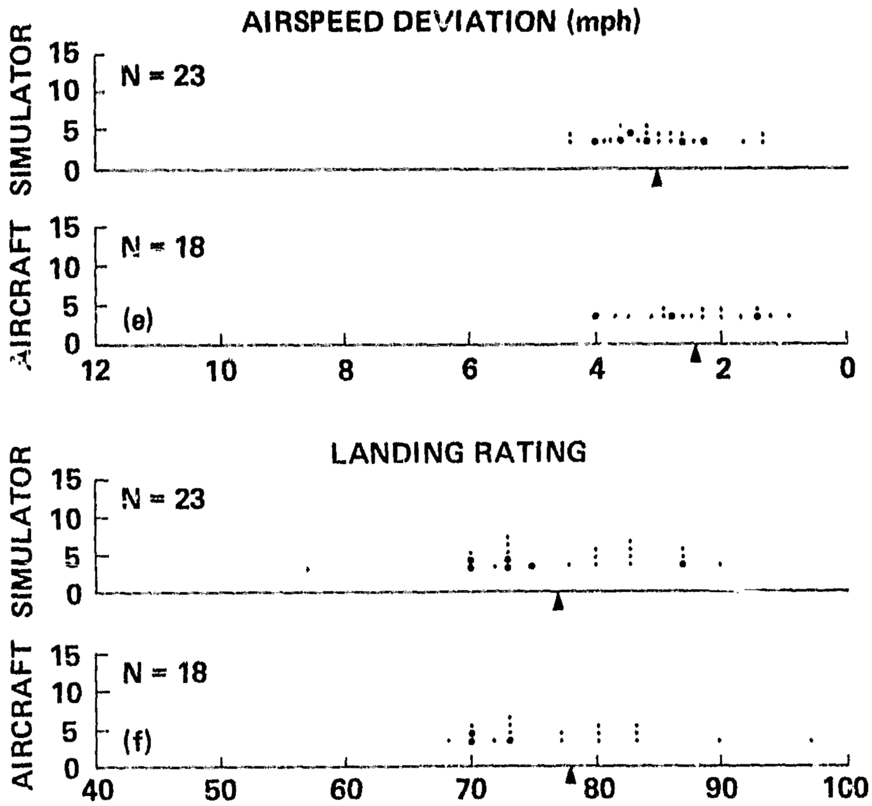

Figure 7.- Continued.

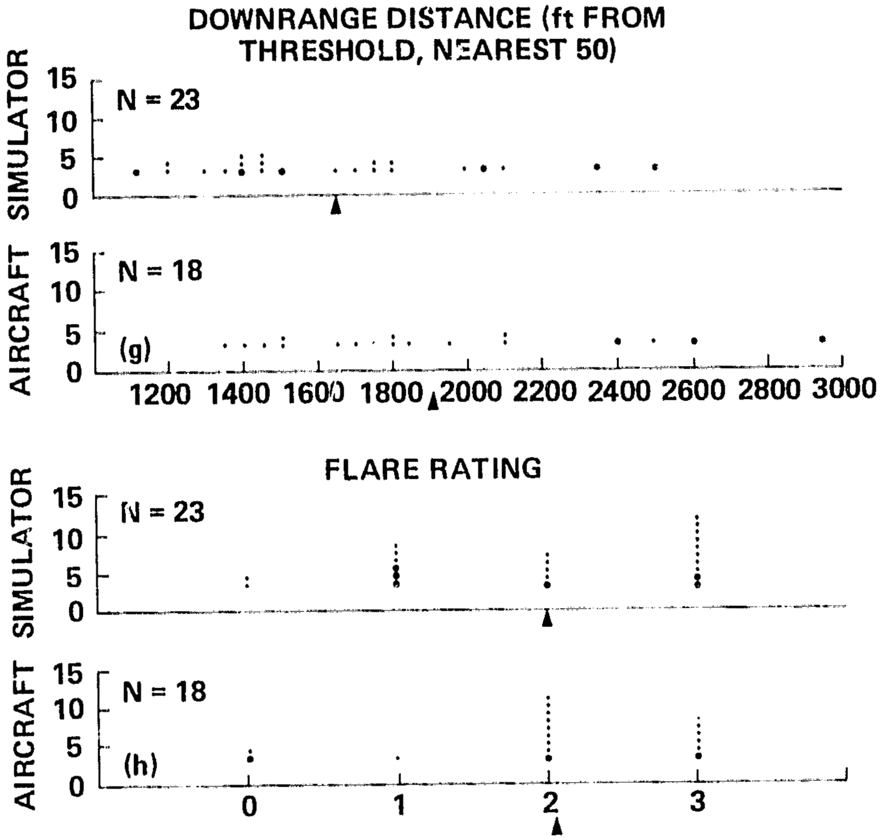

Figure 7.- Concluded.

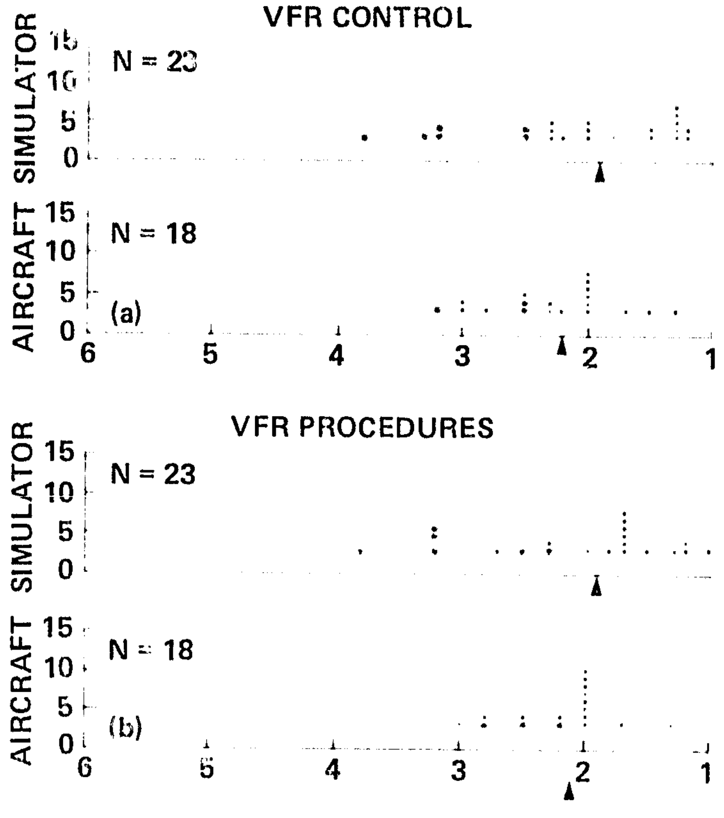

Figure 8.- Phase II check-pilot scores for DC-10 trainees.

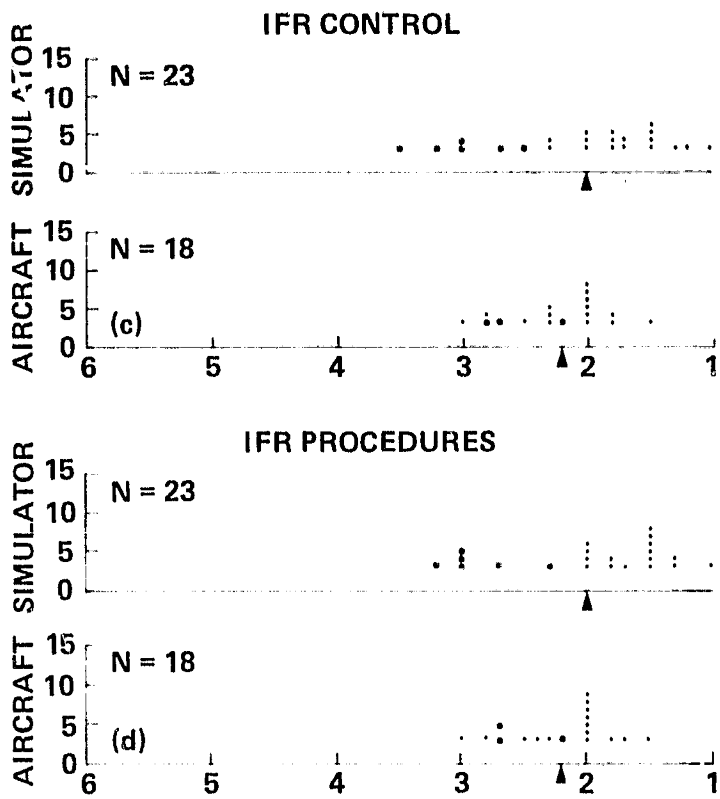

Figure 8.– Concluded.

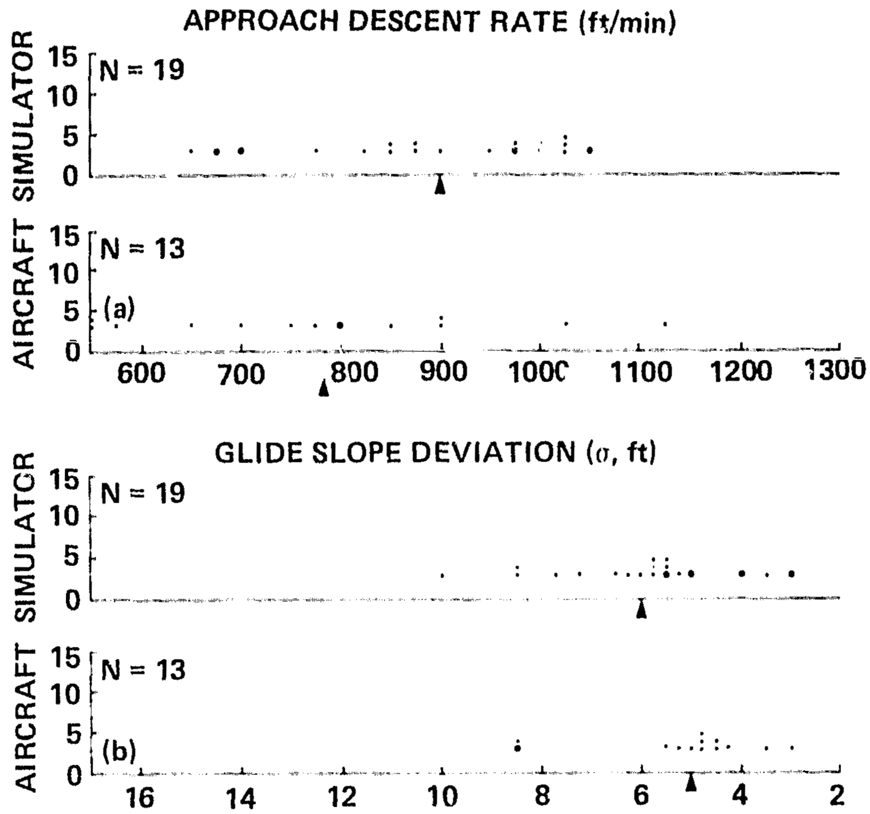

Figure 9.- Phase II aircraft-measured data for DC-10 trainees.

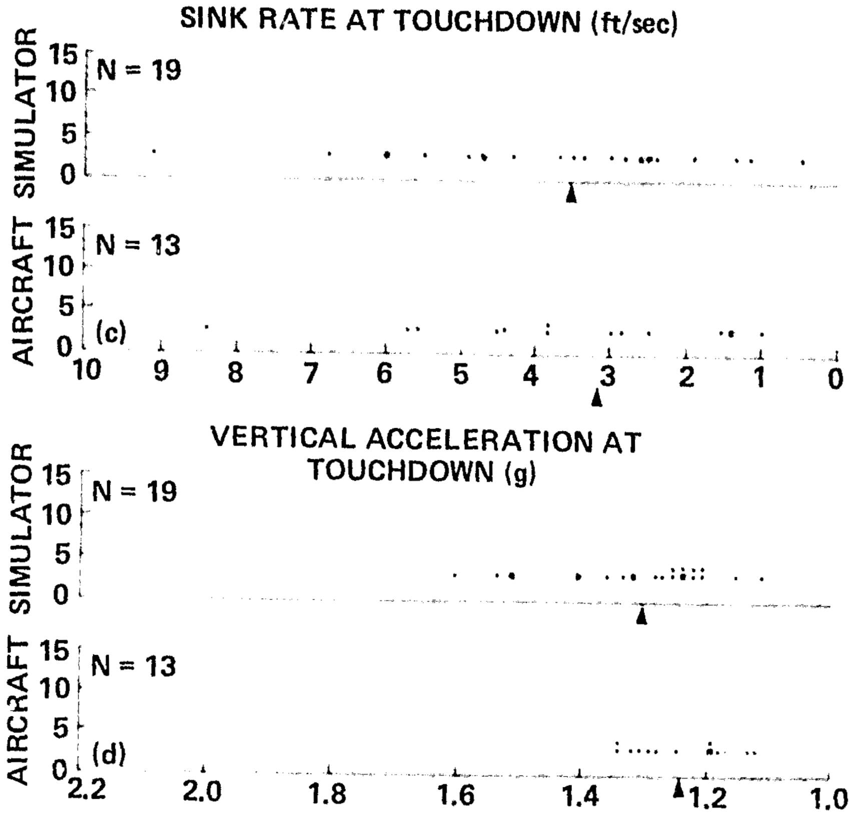

Figure 9.- Continued.

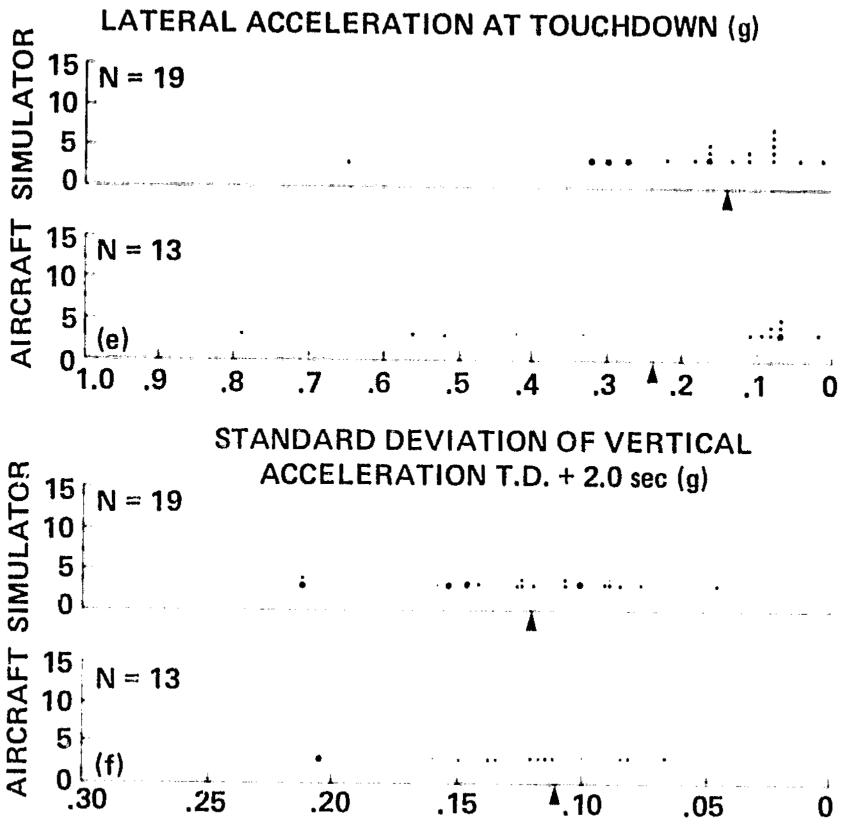

Figure 9.- Continued.

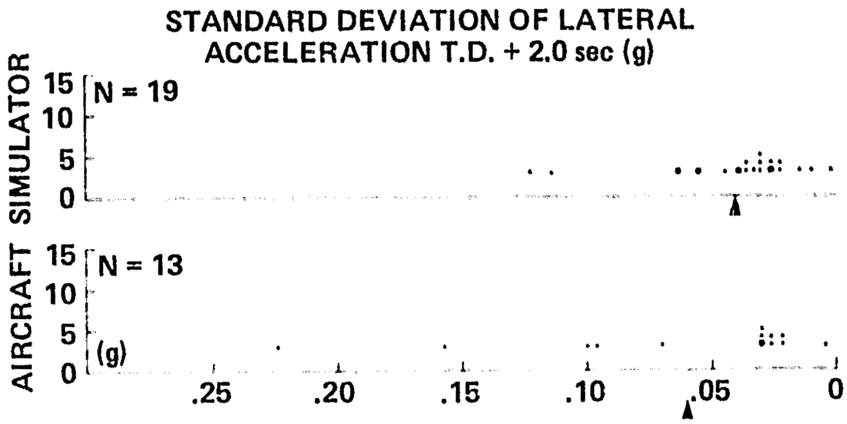

Figure 9.- Concluded.

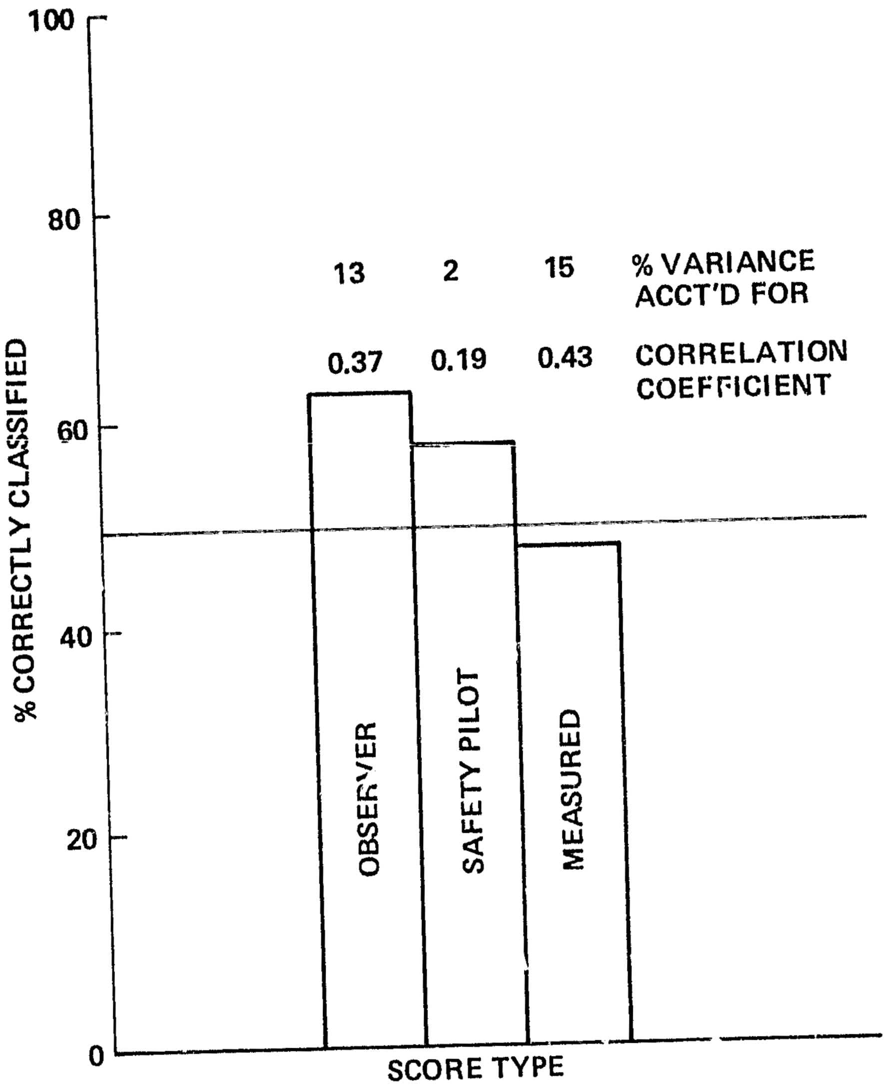

Figure 10.- Percent B-727 trainees correctly classified as having had aircraft or simulator training.

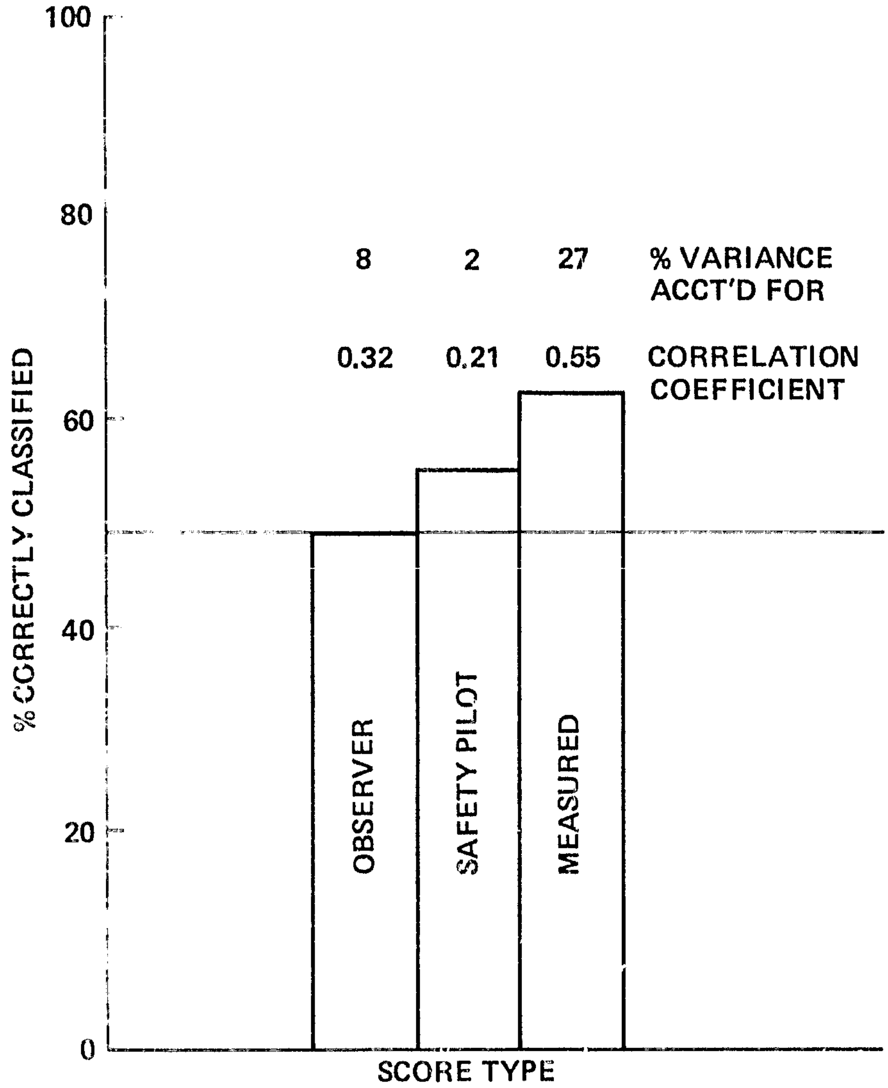

Figure 11.- Percent DC-10 trainees in Phase I correctly classified as having had aircraft or simulator training.

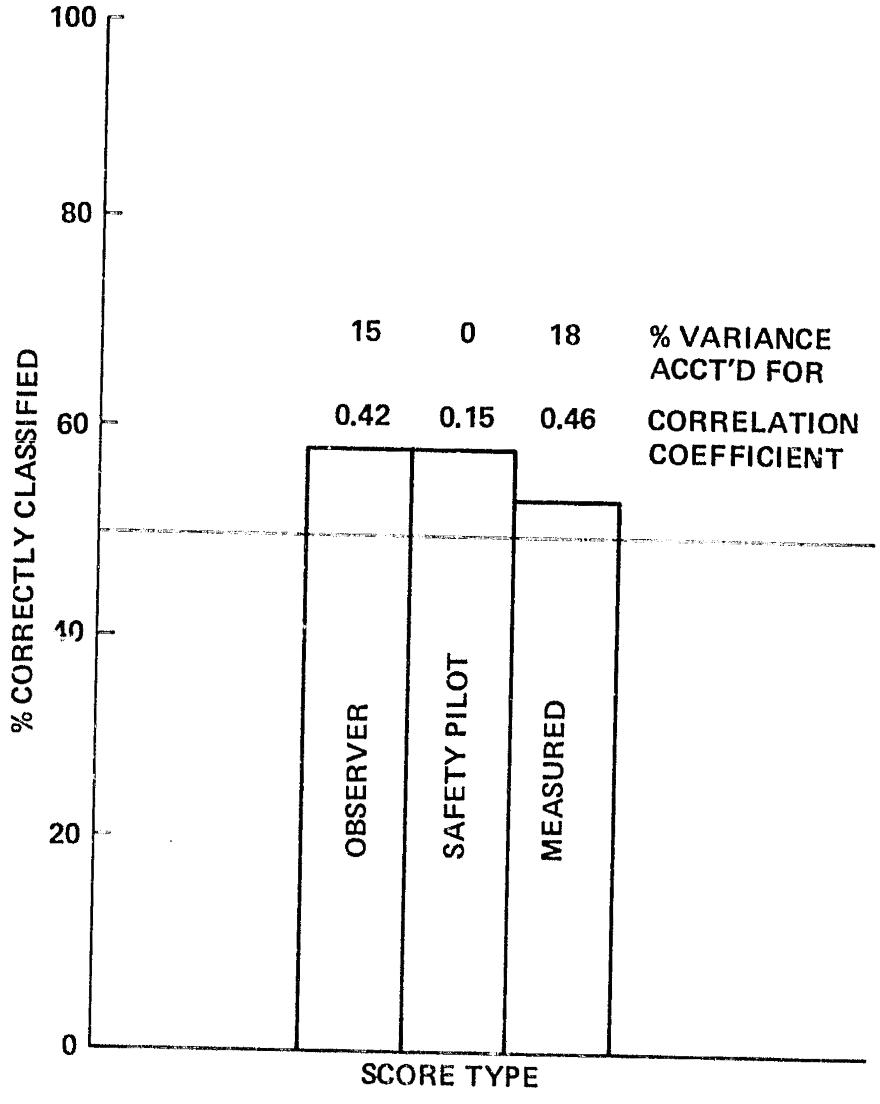

Figure 12.- Percent DC-10 trainees in Phase II correctly classified as having had aircraft or simulator training.

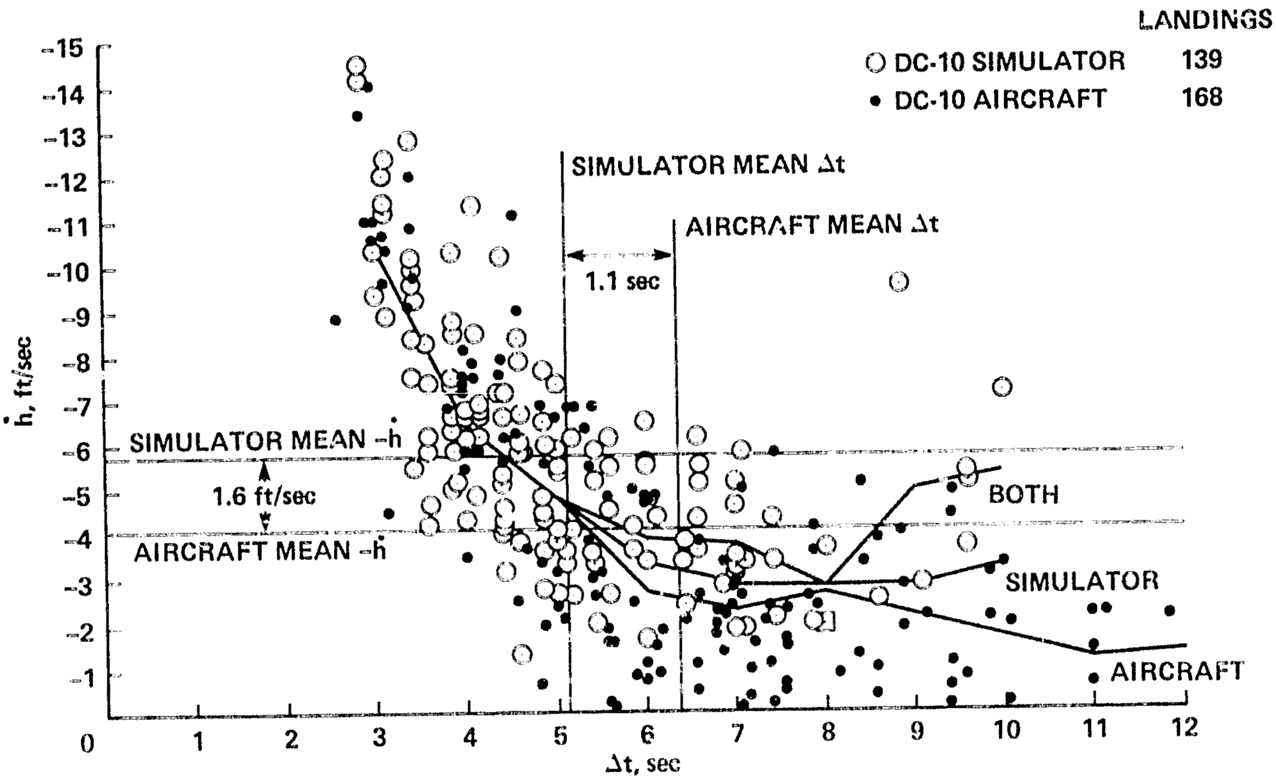

Figure 13.- Relationship between sink rate (ḣ) and timer interval (Δt) from flare initiation to touchdown.

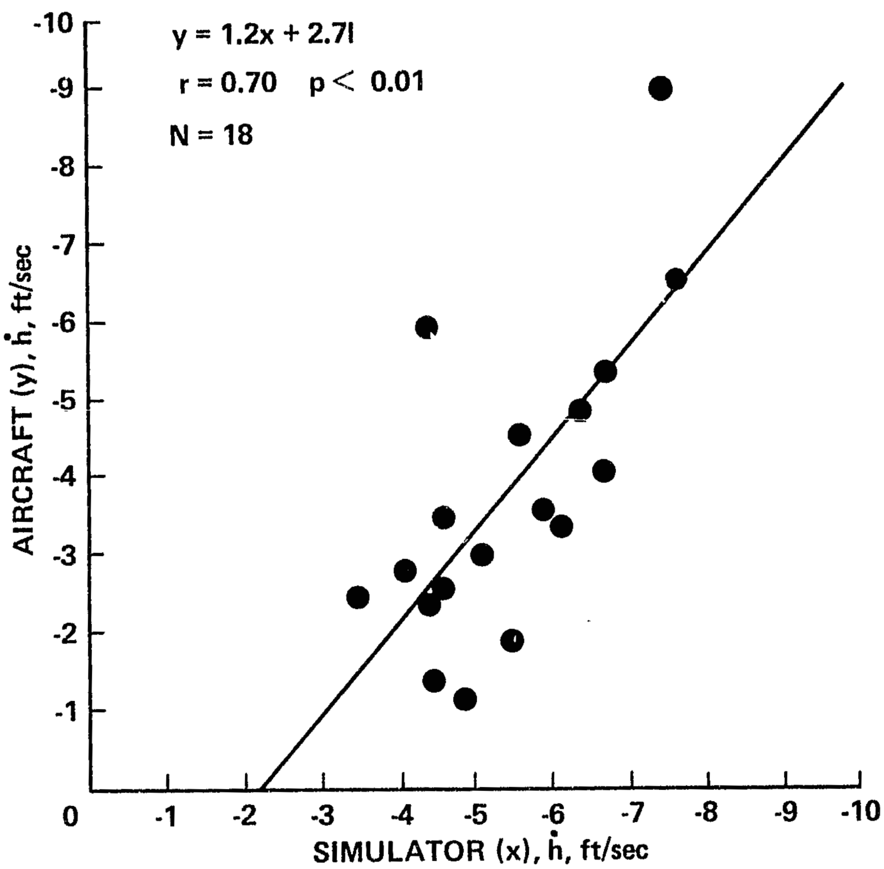

Figure 14.- Relationship between touchdown sink rates in the simulator and aircraft.