THE EFFECTIVENESS OF DIRECTED ATTENTION TO INSTRUMENTS AS A TRAINING AID

by

R. Y. Walker
S. Wapner
D. Bakan
E. S. Ewart

A report on research conducted at the Institute of Aviation Psychology, University of Tennessee, Knoxville, Tennessee, in co-operation with the University of Rochester, Rochester, New York, and the University of Pennsylvania, Philadelphia, Pennsylvania, under the auspices of the National Research Council Committee on Selection and Training of Aircraft Pilots, from funds provided by the Civil Aeronautics Administration and the Tennessee State Bureau of Aeronautics.

October 1946

CIVIL AERONAUTICS ADMINISTRATION
Division of Research
Report No. 69
Washington, D. C.

National Research Council

Committee on Selection and Training of Aircraft Pilots

Executive Subcommittee

M. S. Viteles, Chairman

C. W. Bray	A. I. Hallowell
D. R. Brimhall	J. L. Holland
P. M. Fitts	E. L. Kelly
J. C. Flanagan	P. J. Rulon

LETTER OF TRANSMITTAL

NATIONAL RESEARCH COUNCIL

2101 Constitution Avenue, Washington, D. C.
Division of Anthropology and Psychology

Committee on Selection and Training of Aircraft Pilots

October 31, 1946

Dr. Dean R. Brimhall
Assistant to the Administrator
for Research
Civil Aeronautics Administration
Room 3895, Commerce Building
Washington 25, D. C.

Dear Dr. Brimhall:

Attached is a report entitled The Effectiveness of Directed Attention to Instruments as a Training Aid by R. Y. Walker, S. Wapner, D. Bakan, and E. S. Ewart, submitted by the Committee on Selection and Training of Aircraft Pilots with the recommendation that it be included in the series of Technical Reports of the Division of Research, Civil Aeronautics Administration.

The report is one in the series growing out of research conducted at the Institute of Aviation Psychology, Knoxville, Tennessee, through the cooperation of the University of Tennessee, the State Bureau of Aeronautics, the Committee on Selection and Training of Aircraft Pilots, and the Civil Aeronautics Administration. The study is of particular interest in representing an effort to obtain factual data as a basis for considering differences of opinion among flight instructors with respect to the use of instruments in the early phases of elementary flight instruction. While the investigation provides no definitive answer to the question, it nevertheless does provide certain experimental findings and observations on experimental methods which furnish the basis for further investigations in this area.

Cordially yours,

Morris S. Viteles, Chairman
Committee on Selection and
Training of Aircraft Pilots
National Research Council

MSV:rm

EDITORIAL FOREWORD

The investigation of the effectiveness of directed attention to instruments as a training aid described in this report is one in a series, directed toward the improvement of flight instruction.[1] It was stimulated by differences of opinion among experienced flight instructors regarding the desirability of reference to flight instruments in primary flight instruction.

One point of view is that, in teaching the student pilot to interpret kinesthetic, auditory, and visual cues as to attitude and performance of the plane (e.g., whether or not it is "slipping," "skidding," or gliding too fast or too slow), it is desirable that he associate these sensory cues with what the instruments tell him the plane is doing, until the association is established and reference to the instruments can be dispensed with. The opposing point of view is that such reference to instruments in primary flight training results in "mechanical" flying and teaches the student pilot to continue use of the instruments as "crutches" after reaching the stage where he should not find it necessary to depend on them.

Instrumental aids investigated in the present study include the ball bank, airspeed indicator, and altimeter, with particular reference to the effect of the use of these instruments on student pilot performance in the Straight Climb and Turn maneuvers. One limitation of the investigation is an inadequacy in the experimental design which made it impossible to control simultaneously, in the statistical analysis, the four major sources of variation affecting the criterion scores of the subjects. However, despite this limitation the present study has supplied pertinent, although not definitive, information bearing upon the primary question. Moreover, this investigation served to clarify a number of methodological considerations which will be of importance in the design of further studies of this type.

---

[1] Other studies in this area are described in the following reports:

Kelly, E. Lowell, and Ewart, E. The effectiveness of "Patter" and of "Fundamentals of Basic Flight Maneuvers" as training aids. Washington, D. C.: CAA Division of Research, Report No. 6, December 1942.

Tiffin, Joseph, and Bromer, John. Analysis of eye fixations and patterns of eye movement in landing a Piper Cub J-3 Airplane. Washington, D. C.: CAA Division of Research, Report No. 10, February 1943.

National Research Council Committee on Selection and Training of Aircraft Pilots. The psychology of learning in relation to flight. Washington, D. C.: CAA Division of Research, Report No. 16, June 1943.

Viteles, M. S., et al. A course in training methods for pilot instructors. Washington, D. C.: CAA Division of Research, Report No. 20, September 1943.

Kelly, E. Lowell. The flight instructor's vocabulary. Washington, D. C.: CAA Division of Research, Report No. 22, October 1943.

Ewart, Edwin S., Thompson, Albert S., and Viteles, Morris S. Evaluation of instructional techniques described as effective by flight instructors. Washington, D. C.: CAA Division of Research, Report No. 63, June 1946.

The data for this study were collected in the course of research at the Institute of Aviation Psychology, University of Tennessee, Knoxville, Tennessee, under the supervision of R. Y. Walker, Director. Methods of analysis of photographic records of flight performance are largely the outgrowth of work done by M. S. Viteles, A. S. Thompson, and E. S. Ewart at the University of Pennsylvania, Philadelphia. The statistical analysis was conducted by Seymour Wapner and David Bakan, at the Statistical Office of the Committee, University of Rochester, Rochester, New York. The final report was prepared through the cooperation of the Staff of the Statistical Office and the Editorial Staff of the Committee on Selection and Training of Aircraft Pilots.

CONTENTS

SUMMARY

This experiment which employed as subjects student pilots in the 2nd and 3rd flight classes at the Institute of Aviation Psychology, was conducted to investigate the effect of training student pilots by the method of directed attention to instruments. One group of students, the control group, was given the usual type of flight training. Another group of students, the experimental group, was given special training in which reference to the ball bank, airspeed indicator, and altimeter was emphasized during instruction in the Straight Climb, and Medium Turn maneuvers (180° Turn, 45° Bank).

Various criterion measures of the flight performance of these two groups of students were collected at intervals during the training of each class of students. These measures taken on the control and experimental students, respectively, were subjected to statistical analysis. The method of analysis of variance was used in comparing the performance levels on criteria derived from the motion photographs taken of a hidden instrument panel during each check flight. Analysis of variance was also used in studying the grades assigned by the instructors on the pre-check flights. Analysis of variance was not used in the investigation of the grades assigned by inspectors during the check flights which they administered at intervals during the course, since these data did not warrant such rigorous analysis. Instead, learning curves based on mean grades of the two respective groups of students at each check flight were drawn and compared by inspection.

Because of the nature of the experimental design, it was not possible in the analysis of the photographic data and the instructors' grades to control simultaneously the four variables: methods, classes, instructors, and check flights. Therefore, wherever possible, two different analyses were run on each criterion measure. For one analysis it was assumed that there were no true class differences and the variables "methods," "instructors," and "check flights" were treated simultaneously. For the other analysis it was assumed that there were no true instructor differences, and the variables "methods," "classes," and "check flights" were treated simultaneously. Although it was recognized that neither of these assumptions might be justified, this seemed the most expedient approach since the four variables could not, in any event, be dealt with in a single analysis.

In regard to the major issue of the superiority of one or the other instructional procedure, for the bulk of the criterion variables no statistically significant differences in performance between experimental and control groups were evident. The few statistically significant differences were all in favor of the control group and in terms of most of the remaining variables the control group showed somewhat superior performance, although the differences were not of sufficient size to be statistically significant.

Similarly, the learning curves drawn on the basis of the grades assigned by the inspector of the second class tended to show a consistent superiority on the part of the control group. The learning curves for the grades assigned by the inspector of the third class were ambiguous,

showing no clear-cut superiority for either group. There is evidence to indicate, however, that the grades assigned by the inspector of the third class were less dependable than the grades assigned by the inspector of the third class.

The evidence provided by these analyses indicates that the method of directing attention to instruments as it was applied in this study, is no more effective, and possibly slightly less effective, as a training technique than the more usual training methods. These findings, however, are rendered somewhat equivocal by the fact that no evidence is available regarding the thoroughness with which the instructors of the experimental group employed the experimental training procedure.

In addition to the principal results, analysis of the data yielded a number of supplementary findings having valuable implications in regard to the design of future investigations of this type.

THE EFFECTIVENESS OF DIRECTED ATTENTION TO INSTRUMENTS AS A TRAINING AID

INTRODUCTION

Under the usual conditions of primary flight training no special and specific emphasis is given to the ball bank, airspeed indicator, and altimeter as training aids in teaching Straight Climb and Medium Turn maneuvers. Some experts have suggested that the student's performance of these maneuvers and the facility with which the student becomes proficient in their execution, might be improved by giving greater emphasis to these instruments during training. Such procedures might entail pointing out the relationship between performance of the plane as indicated by these instruments and concomitant sensory cues, e.g., auditory and kinesthetic. The thought in this connection is that by such training the student's recognition of sensory cues, not dependent on instruments, will be speeded and he will learn more rapidly to sense correct climbing speed for example, by auditory cues and will more readily learn, for example, to recognize the kinesthetic cues which indicate that the plane is slipping or skidding.

Other experts do not favor this employment of instruments as training aids. The argument of the adherents to this point of view is that if instruments are used in this manner during training the student will become overdependent on them in spite of warnings to the contrary and will not develop the proper sensitivity to other cues so necessary for good performance. Regarding the ball bank, for example, the proponent of this latter point of view would observe that nearly all students can minimize slipping and skidding by reference to this instrument, but would add that the crucial test of a good pilot is his performance when his attention is directed elsewhere, e.g., outside of the plane.

It was the purpose of this experiment to compare the usual method of flight training, during which no particular emphasis is given to these instruments, with training during which reference to the ball bank, airspeed indicator, and altimeter is regularly and consistently employed as a training aid.

SOURCE OF DATA

The experiment was carried out on the first three flight classes of the Institute of Aviation Psychology.[1] However only the data for the 2nd

---

[1]The Institute of Aviation Psychology was sponsored jointly by the Civil Aeronautics Administration and the Tennessee Bureau of Aeronautics, under the supervision of the National Research Council Committee on Selection and Training of Aircraft Pilots. At the Institute subjects were organized into classes, each student pilot receiving 80 hours of ground school instruction at night and approximately 36 hours of flight instruction. Flight and ground instruction covered daily assignments. Flight and ground school curricula, with the exception of the experimental procedures, followed those recommended by the Civil Aeronautics Administration.

and 3rd flight classes were used in the experiment. One portion of the experiment involving the first class constituted a preliminary experiment serving largely as a trial run. Furthermore, the first class was held at a different airport and two different types of planes were used. The students of the second and third classes were trained on Piper (Cub) J-4 airplanes. All check flights during which criterion data were taken were conducted in a single plane, also a Piper (Cub) J-3. The criterion plane was equipped with a hidden instrument panel. Motion photographic records taken of this panel during the execution of certain maneuvers provided criterion data.

EXPERIMENTAL PROCEDURES

The students of the 2nd and 3rd flight classes were divided into two groups, a control group and an experimental group. The control group received the usual type of training. The experimental group was given special training in which the specified instruments were used as training aids during instruction in Straight Climb and Right and Left Medium Turns as described following:

Experimental Group. The instructors were given indoctrination in training procedures which involved giving special emphasis to the ball bank, airspeed indicator, and altimeter in the instruction of students in the experimental group.

The attention of the students was directed to the airspeed indicator during the Straight Climb, and to the ball bank and altimeter during the Medium Turns. These instruments were available for observation by the students on all flights, dual and solo, up to but not including the 8th check flight (35th hour of flight training). On any dual flight, however, the instructor could prevent the student from viewing the instruments by dropping a cover over them. Occasionally, the instructor would follow this procedure in an attempt to prevent the students from becoming overdependent on these instruments.

The following procedure was outlined in the operational directives for the prosecution of the experiment. "During training procedures (for experimental students) the instructor shall limit special emphasis upon instruments to the two maneuvers selected. During turns he should specifically call the student's attention to the location of the ball on the entry and recovery, and at approximately every 90 degrees of turn during the turn itself. In a similar fashion he should call the student's attention to the altimeter reading prior to the entry, immediately after the entry, prior to the recovery, and immediately after the recovery, and at every 90 degrees of turn during the turn. In the Straight Climb maneuver the instructor should call the student's attention to the airspeed reading while in a level flight position prior to entry, immediately after the entry to the climb, at ten-second intervals during the climb, immediately before the recovery, and immediately after the recovery when again in level flight. The instructor should arrange on any desired flight so that by dropping a cover, these instruments can be viewed only by him. This particular variation should be done on the respective flights which will be half-way in

flying time -- between two successive check flights. This will make it possible for the instructor to train the student to avoid using these instruments as 'crutches.'"[2]

Control Group. The students in the control group received the usual type of flight training prescribed for an approved school. The instructors were told not to give any unusual emphasis to the instruments and to follow the usual procedures prescribed by the regular CAA approved curriculum. The ball-bank instrument was covered during all the instructional flights.

Check Flights and Pre-Check Flights. Each student pilot was given eight check flights by a CAA inspector, each check flight being represented by a "standard flight" (i.e., a specified series of maneuvers in a standard sequence as indicated in Table 1, during which the various criterion measures were obtained. These check flights were administered during the periods of training at 5, 10, 15, 20, 25, 30, 34, and 35 hours. All check flights except the seventh and eighth (after 34 and 35 hours of training, respectively) were administered under the instrument-no instrument conditions under which the subjects had been trained, the check flights administered to the control group being given with the ball-bank indicator covered and to the experimental group with all instruments visible. The seventh check flight was administered to all subjects (both experimental and control) with all instruments visible, whereas the eighth check flight was administered to all subjects (both experimental and control) with the ball-bank indicator and airspeed indicator covered. It was felt desirable to conduct two respective check flights during which all subjects, irrespective of how they were trained, would be tested under conditions first of "instruments visible," and second, of "instruments not visible."

In addition to the check flights administered by a CAA inspector, grades were given to each subject by his flight instructor on the "pre-check flight" lesson, i.e., on performance during the instructional period just preceding each check flight. During these pre-check flight periods, the instructor graded the subject on his performance of every maneuver included in the check flight to follow, although the sequence in which maneuvers were performed during the pre-check flight lesson period was purposely altered from the order in which they appeared in the standard flight employed in the check flight to follow.

In Table 1 are presented the maneuvers included in the various check flights. The order of maneuvers in Table 1 represents the order in which they appeared in the various standard flights. As noted above, the order in which the maneuvers were presented in the pre-check flight lesson period was not the same as the sequence employed during the check flight proper.

---

[2]Although preliminary training of instructors was carried out in terms of this directive, there are no data available indicating the degree to which this indoctrination of instructors was successful, or the degree to which training of the "experimental" students differed from the training of the "control" subjects.

TABLE 1

MANEUVER NUMBERS AND CHECK FLIGHT SEQUENCE

Man. No.	Maneuver	Maneuvers in Check Flight							
		1 5 hr.	2 10 hr.	3 15 hr.	4 20 hr.	5 25 hr.	6 30 hr.	7 34 hr.	8 35 hr.
1.	Taxi	x	x	x	x	x	x	x	x
2.	Take-off	x	x*	x	x*	x	x	x*	x*
3.	Straight and Level	x	x*	x	x*	x	x	x*	x'
4.	S-Turns		x						
5.	Shallow 8 around pylon			x	x				
6.	Steep 8 around pylon					x	x	x	x
7.	Straight Climb**	x	x'	x	x'	x	x	x'	x'
8.	L. 90° Climbing turn	x	x*	x	x*	x	x	x*	x'
9.	R. 90° Climbing turn								
10.	L. 90° Turn, 15° Bank					x	x	x	x
11.	R. 90° Turn, 15° Bank					x	x	x	x
12.	L. 90° Turn, 30° Bank	x	x						
13.	R. 90° Turn, 30° Bank	x	x						
14.	L. 180° Turn, 45° Bank		x*	x	x*	x	x	x*	x'
15.	R. 180° Turn, 45° Bank		x*	x	x*	x	x	x'	x'
16.	L. 360° Turn, 45° Bank								
17.	R. 360° Turn, 45° Bank								
18.	L. 360° Turn, Steep			x	x*	x	x	x*	x'
19.	R. 360° Turn, Steep			x	x*	x	x	x*	x'
20.	Normal Power-off Stall	x	x*	x	x*	x	x	x*	
21.	Normal Power-on Stall								
22.	Complete Power-off Stall								
23.	Complete Power-on Stall								
24.	One-turn Spin						x		
25.	One-turn Spin						x		
26.	1080 Spiral						x		
27.	1080 Spiral								
28.	Straight Glide	x	x*	x	x*	x	x	x*	
29.	90° L. Gliding Turn	x	x*	x	x*	x	x	x*	
30.	90° R. Gliding Turn								
31.	Forward Slip		x	x	x	x	x	x	x
32.	Landing (not spot)		x*						
33.	Spot Landing			x	x*	x	x	x*	

*Indicates maneuver is photographed.

**Maneuvers 7, 14, and 15 are the experimental maneuvers on which the experimental students received special training.

SUBJECTS

The subjects in this study were drawn from the University of Tennessee and the area of Knoxville, Tennessee. The ages of the students ranged from 18 to 38. Both male and female students participated in the experiment, although approximately 80 per cent of the subjects were male. The 2nd flight class had 14 experimental students and 9 control students. The 3rd flight class had 10 experimental and 14 control students. An attempt was made to match the students of the control and experimental groups in terms of age, sex, and scores on selection tests. The selection tests included the Test of Mental Ability, the Test of Mechanical Comprehension, the Personal History Inventory, the Test of Aviation Information, the Desire to Fly Inventory, the Two-Hand Coordination Test, and the Mashburn Serial Coordination Test.

CRITERIA

The performance levels of the students at specified points throughout the course were evaluated by the use of several criterion measures of flight performance. There were three sources for these criterion measures:

1. The grades assigned by the instructor on each pre-check flight.[3] These were grades on particular maneuvers and grades on the flight as a whole.

2. Motion photographs taken on the 2nd, 4th, 7th, and 8th check flights. These photographs provided records of readings of various instruments on a hidden instrument panel.[4]

3. Measures obtained by the inspector on each check flight. During the check flights the inspectors entered their observations regarding the student's flight performance on the Ohio State Flight Inventory.[5] On the basis of these

---

[3]The term check flight will be used synonymously with pre-check flight in the text. However, it should be indicated that all instructor-assigned grades were based on performances during the pre-check flight period, whereas criterion measures based on the photographic records and on inspector grades were based on the check flight performance.

[4]A more complete description and discussion of the photographic installation is presented in: Viteles, M. S., and Thompson, A. S. An analysis of photographic records of aircraft pilot performance. Washington, D. C.: CAA Division of Research, Report No. 31, July 1944. A detailed description of the record reading procedures employed in this study are on file in the offices of the Committee on Selection and Training of Aircraft Pilots.

[5]This criterion instrument which comprises a series of check sheets by means of which critical details of the plane's performance during each maneuver of the check flight are recorded by the check pilot and later evaluated, is described in detail in: Edgerton, H. A., and Walker, R. Y. History and development of the Ohio State Flight Inventory, Part I: Early versions and basic research. Washington, D. C.: CAA Division of Research, Report No. 47, July 1945. NRC Committee on Selection and Training of Aircraft Pilots. History and development of the Ohio State Flight Inventory, Part II: Recent versions and current applications. Washington, D. C.: CAA Division of Research, Report No. 51, November 1945.

entries an over-all score from the Inventory was derived by summating (with the addition of a constant) the positive unit weights assigned those items representing satisfactory performance. Furthermore, on the basis of his own evaluation of the flight performance, the inspector also assigned grades (a) to each individual maneuver executed (referred to hereafter as "maneuver grades") and (b) to the students' performance on the flight as a whole, (referred to hereafter as "over-all grade").

STATISTICAL ANALYSIS

Allocation of Subjects. Table 2 indicates the number of control and experimental students assigned to each instructor in each class.

TABLE 2

ALLOCATION OF CASES

Instructor	2nd Flight Class		3rd Flight Class	
	Exp.	Control	Exp.	Control
1	x	6	6	x
2	x	3	1	x
3	4	1	x	4
4	6	x	x	5
5	5	x	x	6
6	x	x	4	x
N	14	9	10	14

Several characteristics of the allocation of cases are noteworthy:

1. Each instructor, with the exception of Instructor 6, had both control and experimental students. Because Instructor 6 did not have any control students it was decided to drop his four students from the analysis.

2. With the exception of Instructor 3, the instructors had only control or experimental students in each class; and, the instructors who had control students in one class had experimental students in the other class and vice versa.

Variables Affecting Criterion Measures. In addition to the major variable, the method of training, there are three other variables in the experiment which may have modified the performance levels and criterion scores of the students. These variables are the classes, the check flights and pre-check flights, and the instructors. In addition to the effect that differing teaching practices of the various instructors might have had on the performance levels of the students, the different grading practices of the instructors could also have produced differences in the instructor criterion

scores. Differences between classes may have occurred because of differences in the weather conditions under which the students took their flight training. Differences in the criterion scores from check flight to check flight may be anticipated because of learning. However, due to the fact that, in general, in a given flight class individual instructors had either experimental or control students, but did not have both experimental and control students, class variance and instructor variance were confounded and it was not possible to treat these variables simultaneously. The implications of this situation are presented and discussed in detail in Appendix 1.

Analyses Employed: Analysis of Variance. It is unfortunate that due to limitations in the experimental design the four variables involved in the experiment, methods, classes, instructors, and check flights, could not be evaluated simultaneously. Since the major concern of the experiment was the methods variable, and since differences from check flight to check flight could be expected, two types of analysis could be done:

1. It could be assumed that there are no true class differences and hence the analyses could be carried out taking account of the variables methods, instructors, and check flights.

2. It could be assumed that there are no true instructor differences, and the analyses could be carried out taking account of the variables methods, classes, and check flights.

Since class and instructor variance could not be treated simultaneously both of the above analyses were carried out. If both of these statistical analyses yield similar results with respect to methods, sufficient evidence would be supplied to warrant conclusions to be drawn even though the experimental design has these limitations.

Analysis of variance was chosen as the statistical technique for analysing the data because it permits control of several variables, increases the precision of the experiment, and permits the evaluation of interactions. The two types of analysis which were carried out are presented diagrammatically in Figures 1 and 2.

Each axis of each of the figures represents a major variable. In one type of analysis (Figure 1), the major variables dealt with are methods, instructors, and check flights. In the second type of analysis (Figure 2), the major variables are methods, classes, and check flights. The entry in each cell is represented by the mean of the criterion scores for each group of students characterized by the variables of each of the axes.

A separate analysis of variance was done for each of the criterion measures presented in Tables 3 and 4.

F-tests were run for each of the major variables and each first-order interaction for each criterion measure. The error term used for the evaluation of the first-order interactions was the second-order interaction variance. The major variables were evaluated by comparing their variances with a pooled error variance consisting of the first-order interactions (which were not significantly greater than the second-order interaction)

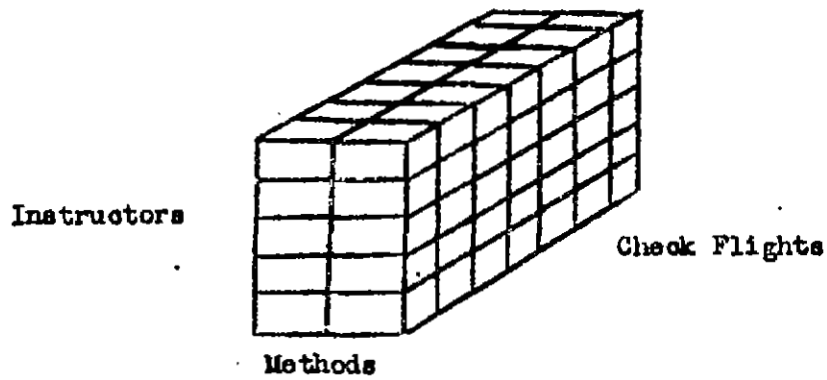

FIGURE 1

ANALYSIS FOR VARIABLES: METHODS, INSTRUCTORS, AND CHECK FLIGHTS

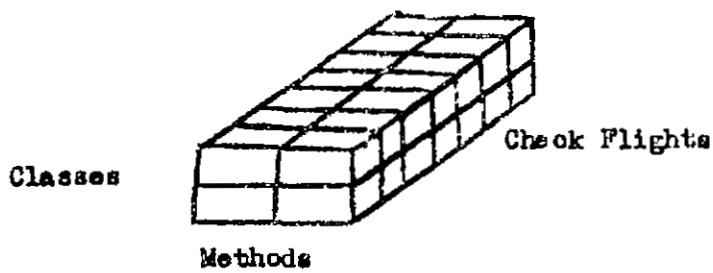

FIGURE 2

ANALYSIS FOR VARIABLES: METHODS, CLASSES, AND CHECK FLIGHTS

and the second-order interaction. The pooled error variance was determined by adding the appropriate sums of squares and dividing by the sum of the appropriate degrees of freedom. For the analyses involving methods, instructors, and check flights there were seventeen applications involving the nine criterion measures for which data were available, as indicated in Table 3. A separate analysis of variance was carried out for each of these seventeen applications. For the analysis involving methods, classes, and check flights, there were available data for 22 applications of twelve criterion measures as indicated in Table 4. A separate analysis of variance was carried out for each of these applications of criterion measures.[6]

---

[6]The differences in the number of criterion measures for which data were available from one type of analysis to another result from the fact that the number of cells in terms of which the data are categorized for each of the two types of analysis is not the same. Since the first type of analysis (methods, instructors, and check flights) involves a greater number of cells to be filled, it is to be expected that there would be a greater opportunity for cells to be unfilled. In the instance where too many cells had missing data the analysis was not carried out.

It should also be noted that in the breakdown by instructors the data from the photographic records from the second check flight were in general too sparse to warrant inclusion of data from this check flight in the analysis of "methods, instructors, and check flights."

TABLE 3

CRITERION MEASURES IN ANALYSIS INVOLVING METHODS, INSTRUCTORS, CHECK FLIGHTS

	Maneuver		Criterion Measure	
7.	Straight Climb	1.	Airspeed variation	(Phot.)
		2.	Maneuver grade	(Instr.)
14.	Left 180° Turn, 45° Bank	1.	Altitude variation - turn	(Phot.)
		2.	Altitude difference - recovery	(Phot.)
		3.	Slips - turn	(Phot.)
		4.	Slips - recovery	(Phot.)
		5.	Skids - turn	(Phot.)
		6.	Skids - recovery	(Phot.)
		7.	Maneuver grade	(Instr.)
15.	Right 180° Turn, 45° Bank	1.	Altitude variation - turn	(Phot.)
		2.	Altitude difference - recovery	(Phot.)
		3.	Slips - turn	(Phot.)
		4.	Slips - recovery	(Phot.)
		5.	Skids - turn	(Phot.)
		6.	Skids - recovery	(Phot.)
		7.	Maneuver grade	(Instr.)
	All maneuvers	1.	Over-all flight grade	(Instr.)

TABLE 4

CRITERION MEASURES IN ANALYSIS INVOLVING METHODS, CLASSES, CHECK FLIGHTS

	Maneuver		Criterion Measure
7.	Straight Climb	1.	Airspeed variation (Phot.)
		2.	Maneuver grade (Instr.)
14.	Left 180° Turn, 45° Bank	1.	Altitude difference - entry (Phot.)
		2.	Altitude variation - turn (Phot.)
		3.	Altitude difference - recovery (Phot.)
		4.	Slips - entry (Phot.)
		5.	Slips - turn (Phot.)
		6.	Slips - recovery (Phot.)
		7.	Skids - entry (Phot.)
		8.	Skids - turn (Phot.)
		9.	Skids - recovery (Phot.)
		10.	Maneuver grade (Instr.)
15.	Right 180° Turn, 45° Bank	1.	Altitude difference - entry (Phot.)
		2.	Altitude variation - turn (Phot.)
		3.	Altitude difference - recovery (Phot.)
		4.	Slips - entry (Phot.)
		5.	Slips - turn (Phot.)
		6.	Slips - recovery (Phot.)
		7.	Skids - entry (Phot.)
		8.	Skids - recovery (Phot.)
		9.	Maneuver grade (Instr.)
	All maneuvers	1.	Over-all flight grade (Instr.)

Analyses Employed: Learning Curves. For each of the two flight classes there was a different inspector. This fact makes the analysis of the inspector criterion measures still more difficult, since it adds another variable which is confounded with the class variable. Furthermore, there is evidence that the inspector of the third class was less reliable than the inspector of the second class. The grades assigned by the inspector of the third class also showed much less relationship to the grades assigned by the instructor than the inspector of the second class.[7]

In view of these limitations it was decided not to subject the inspector criterion measures to any refined statistical analysis. However, in order to obtain some over-all view of the results on these criterion measures, the group learning curves were drawn separately for each class (i.e., for each inspector).

RESULTS FOR VARIABLES: METHODS, INSTRUCTORS, AND CHECK FLIGHTS

Tables 5A, 5B, and 5C present a summary of results of the analyses of variance for the seventeen criterion measures on which there were sufficient data to carry out the analyses involving methods, instructors, and check flights. In these tables the significance of the F-tests are given for each evaluation of a major variable and each first-order interaction. This is done for each of the criterion measures. The 5 per cent level of confidence was used in evaluating each F, except in the evaluation of the

---

[7] The two inspectors differed in "reliability" as measured by test-retest correlations on the 7th and 8th check flights.

	2nd Flight Class (Inspector A)	3rd Flight Class (Inspector B)
Over-all Score	.91	.47
Mean Maneuver Fl. Gr.	.95	.71
OSFI	89	.72

Even more striking is the marked disagreement between the inspector's grades and the instructor's grades. The correlation coefficients between the inspector's and the instructor's grades are given below separately for flight class and check flights.

Criterion	Flight Class	Check Flight (Pre-Check Flight) 1	2	3	4	5	6	7
Over-all Score	2	.42	.58	.40	.66	.62	.75	.86
	3	-.11	-.11	-.13	-.10	.44	.29	.07
Mean Maneuver Fl. Gr.	2	.41	.50	.45	.68	.64	.76	.71
	3	.00	-.06	-.19	03	.42	.13	.20

TABLE 5A

SUMMARY OF RESULTS OF ANALYSES OF VARIANCE: METHODS, INSTRUCTORS, AND CHECK FLIGHTS

(Direction of Mean Differences Between Experimental and Control Groups and Significance of F-Tests)

Photographic Records for Maneuver 7, Straight Climb, and Maneuver 14, Left 180° Turn, 45° Bank

	Man. 7			Man. 14			
	Airspeed Variation	Altitude Variation (turn)	Altitude Difference (recov.)	Slips (turn)	Slips (recov.)	Skids (turn)	Skids (recov.)
Methods group showing better performance*	C	C	C	=	C	C	C
Variance							
Methods	n.s.**	n.s.	n.s.	n.s.	n.s.	.01	n.s.
Instructors	.05-.01	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Check flights	n.s.	n.s.	n.s.	<.01	n.s.	n.s.	n.s.
Methods x Instructors	n.s.	n.s.	n.s.	n.s.	n.s.	<.01	n.s.
Methods x Check flights	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Instructors x Check flights	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

*When the experimental group shows better performance it is indicated by E; when the control group shows better performance it is indicated by C; when there is no difference it is indicated by =. Of course, this only shows the direction and does not necessarily mean that the difference is significant.

**Not significant at 10% level of confidence for methods variances. For all other variances n.s. means not significant at the 5% level of confidence.

TABLE 5B

SUMMARY OF RESULTS OF ANALYSES OF VARIANCE: METHODS, INSTRUCTORS, AND CHECK FLIGHTS

(Direction of Mean Differences Between Experimental and Control Groups and Significance of F-Tests)

Photographic Records for Maneuver 15, Right 180° Turn, 45° Bank

	Man. 15					
	Altitude Variation (turn)	Altitude Difference (recov.)	Slips (turn)	Slips (recov.)	Skids (turn)	Skids (recov.)
Methods group showing better performance*	C	E	C	C	E	C
Variance						
Methods	n.s.**	n.s.	<.01	n.s.	n.s.	n.s.
Instructors	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Check flights	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Methods x Instructors	.05-.01	n.s.	.05-.01	n.s.	n.s.	n.s.
Methods x Check flights	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Instructors x Check flights	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

*When the experimental group shows better performance it is indicated by E; when the control group shows better performance it is indicated by C; when there is no difference it is indicated by =. Of course, this only shows the direction and does not necessarily mean that the difference is significant.

**Not significant at 10% level of confidence for methods variances. For all other variances n.s. means not significant at the 5% level of confidence.

TABLE 50

SUMMARY OF RESULTS OF ANALYSES OF VARIANCE: METHODS, INSTRUCTORS, AND CHECK FLIGHTS

(Direction of Mean Differences Between Experimental and Control Groups and Significance of F-Tests)

Instructor Grades for Maneuver 7, Straight Climb; Maneuver 14, L. 180° Turn, 45° Bank; Maneuver 15, R. 180° Turn, 45° Bank; and Over-all Grades

	Man. 7	Man. 14	Man. 15	Over-all Grade
Methods group showing better performance*	C	E	C	C
Variance				
Methods	.10-.05	n.s.**	n.s.	.10-.05
Instructors	<.01	n.s.	<.01	<.01
Pre-check flights	<.01	<.01	<.01	<.01
Methods x Instructors	n.s.	n.s.	.05-.01	n.s.
Methods x Pre-check flights	n.s.	n.s.	n.s.	n.s.
Instructors x Pre-check flights	<.01	n.s.	n.s.	n.s.

*When the experimental group shows better performance it is indicated by E; when the control group shows better performance it is indicated by C; when there is no difference it is indicated by =. Of course, this only shows the direction and does not necessarily mean that the difference is significant.

**Not significant at 10% level of confidence for methods variances. For all other variances n.s. means not significant at the 5% level of confidence.

methods variance in which the 10 per cent level of confidence was used.[8]

If the F is not significant, it is denoted in the table by "n.s." Where the F is significant, the p value is given. These tables also indicate which of the two methods groups show better mean performance. For the instructor measures, the higher the score the better the performance. For the photographic criterion measures, the lower the score the better the performance.

Methods. Examination of these tables shows that the control group has better mean performance than the experimental group for 13 out of the 17 criterion measures considered. In four of the instances in which the control group shows better performance the F's for methods are significant. None of the F's are significant where the experimental group shows better performance.

Significant differences were found at less than the 1 per cent level of confidence for Skids (turn) on Maneuver 14, (Left 180° Turn with 45° Bank), and for Slips (turn) on Maneuver 15, (Right 180° Turn with 45° Bank) the control group being superior in all of these instances. The instructors' maneuver grades for Maneuver 7, Straight Climb, and the instructors' over-all grades are both significant between the 10 per cent and 5 per cent levels of confidence, control subjects again being superior.

It appears, then, that wherever there are significant differences between the methods in these analyses, these differences tend to indicate

---

[8] The level of significance of the "F" value indicates the number of times in 100 that the difference in question could be expected to arise by chance alone. An "F" value significant at the 5 per cent value, for instance, indicates that the difference could be expected to arise by chance only 5 in 100 times. Some justification seems necessary for the use of a level of confidence as high as 10 per cent instead of the orthodox 5 per cent level. This justification comes from a review of the general context of the experiment. In this experiment the major interest is that of determining whether instructors should or should not be urged to direct attention to special instruments in training students on certain maneuvers. If, in this experiment, the experimental group showed better performance than the control group, it would be necessary to be very certain that the difference is significant before a recommendation should be made to change the method of instruction. It would be necessary to be quite certain that there is a true difference because of the effort and cost required to introduce the change. Under these circumstances, it would be advisable to use a very conservative confidence level. However, in the sample studied the control group for the most part showed better performance than the experimental group. With this kind of result it is more conservative to use the 10 per cent level of confidence. The effect of using a fairly liberal confidence level in this instance is to reject the hypotheses of no difference between methods if there is even a small indication that the new method has the effect of impairing the performance level.

that the training method of directed attention to instruments is less effective than the usual training procedure.[9]

Methods x Instructors. If the methods x instructors variance[10] is significant it indicates that the hypothesis that the difference between methods is the same from one instructor to another instructor may be rejected. The analysis indicated that this hypothesis may be rejected, i.e., that there were significant methods x instructor interactions for Skids during the turn in Maneuver 14, and on Maneuver 15 for the photographic measure "Altitude Variation" during the turn, and also for the instructor maneuver grade. It may be inferred that the difference between methods varies significantly depending on the identity of the instructor involved and, therefore, that for these criterion measures differences between methods may be, in part, a function of differences among instructors.[11]

Methods x Check Flights. Inspection of Tables 5A, 5B, and 5C shows that there are no significant methods x check flights interactions. In other words, on the basis of these data the hypothesis that the difference between the methods remains the same from check flight to check flight cannot be rejected.

Instructors. Differences in criterion measures attributable to instructors alone may be a function of a number of factors. In terms of photographic criteria, a difference attributable to instructors may reflect differences in quality of instruction. In terms of instructors' grade, differences attributable to instructors may also result from differences between instructors in grading standards.

Examination of Tables 5A, 5B, and 5C indicates that of the twelve criterion measures taken from the photographic records instructor variance was significant in only one case, i.e., Airspeed Variation during Maneuver 7 (Straight Climb). However, of the criterion measures based on instructor grade, significant instructor variance (at the 1 per cent level of confidence) is evident in three of the four cases. The conclusion seems warranted that differences in grading methods represented a marked source of variance, although in the case of Airspeed Variation in the Straight Climb (Maneuver 7) the significant variance cannot be attributed to differences in grading methods, since it was demonstrated on the basis of a photographic

---

[9]The methods differences obtained could also be a reflection of instructors x classes interactions (see Appendix 1).

[10]Methods x instructors variance accounts for differences in criterion scores which result when the adequacy of teaching or grading procedures used by the instructors is a function of the set of training conditions under which the instructor works.

[11]The methods x instructors interaction obtained could also be a reflection of class differences (see Appendix 1).

criterion.[12]

Check Flights. A significant check flight variance was indicated for only one of the photographic criterion variables, i.e., Slips in Maneuver 14 (180° Turn Left). The mean for check flight 8 was higher, showing poorer performance than the mean for check flight 4 or 7. Since the ball bank was not visible to either control or experimental students during this maneuver it might be indicated that its absence from view was detrimental to performance, at least on this maneuver. However, because this significant variance appeared for the Slip and Skid variables in turns only on this maneuver no definitive conclusions can be drawn, except to note that the very fact that only one significant check flight variance among photographic criteria was evident is, in itself, of importance. However, it should be recognized that photographic data from only the 4th, 7th, and 8th check flights were included in the "methods, instructors, check flights" analysis. These results do not, then, preclude the possibility that significant check flight variances might have been evident for the photographic variables had data from earlier check flights also been included.

All of the instructor criteria show significant check flight variances, as might be expected, since grades throughout the course were assigned in terms of a fixed standard rather than a floating standard, i.e., all performances were judged using the performance expected of a private pilot as a standard. Since there is a progressive increase in the means of these grades from check flight to check flight, the progressive learning from stage to stage is undoubtedly reflected.[13] Since the instructors did not administer an eighth pre-check flight it is impossible to ascertain the effect on instructor grades of covering the ball bank.

Instructors x Check Flights. A significant instructor x check flight interaction was found for only one of the criterion measures, i.e., the maneuver grades assigned by instructors on Maneuver 7 (Straight Climb). This may be interpreted as indicating that the increments in performance between check flights, as indicated by instructor grade, are different for students of different instructors. The fact that this significant interaction appears only for instructor grade may indicate that differences in grading standards or procedures may be a factor. However, the fact that such significant interaction occurred for only one maneuver prevents definitive conclusions being drawn, except to note that there was markedly less evidence of instructor check flight interaction than of variance attributable to instructors. That is, although there was considerable evidence of differences in grading procedures among instructors, there was less evidence of differences among instructors in grading the apparent increments of performance from check flight to check flight.

---

[12]The instructor differences obtained could also be a reflection of methods x classes interactions (see Appendix 1).

[13]It may also be possible that the instructors systematically give higher grades on subsequent check flights than that which accurately reflects the performance level. Both factors may be operative in this instance. (See Discussion, page 29.)

RESULTS FOR VARIABLES: METHODS, CLASSES, AND CHECK FLIGHTS

Tables 6A, 6B, and 6C present a summary of results of the analyses of variance for the 22 criterion measures on which there were sufficient data to carry out the analyses involving methods, classes, and check flights. In these tables, the significance of the F-tests is given for each evaluation of the major variable and each first-order interaction.[14] The tables also indicate which of the two methods groups shows better mean performance. As in the case of the previous analyses, the higher the score for the instructor measures, the better the performance. For the photographic criterion measures, the lower the score, the better the performance.

Methods. Examination of these tables shows that in only one instance, Airspeed Variation on Maneuver 7, is the variance for methods significant.[15] In this instance, the control group shows better mean performance. For the remaining criterion measures the control group shows better mean performance in 9 instances although these are not significant. It should be noted that none of the variables yielding significant variances in the "methods, instructors, and check flights" analyses carried significant F's in the present analysis.

Classes. Class differences which are significant occur on 14 out of 22 criterion measures. This result is quite striking. However, it must be interpreted with caution, in view of the possibility that the assumption on which the analyses is based is not valid. It will be recalled that the statistical design of these analyses involved the assumption that the instructor variable is not a significant factor in modifying the criterion scores. Since it is not at all unlikely that this assumption is false, and should there be a methods x instructors interaction, the class differences observed may be a reflection of this interaction rather than an index of true differences between the two classes.[16] Moreover, although the previous analysis, based on the assumption that there were no true class differences, yielded no marked evidence of significant instructor variance, or significant instructors x methods interaction, the possibility might also be considered that the apparent class differences reflect instructor-class interactions which were impossible to isolate due to the original design of the experiment.

Check Flights. A significant check flight variance was found in one instance on a photographic criterion, i.e., Airspeed Variation in Maneuver 7. It is to be noted that in this analysis of "methods, classes, and check flights" photographic data from the second flight class were included. For

---

[14]The method of evaluation of the F's in these analyses are the same as those used in the foregoing analyses. (See pages 7 and 8.)

[15]The methods difference obtained could be a reflection of an instructors x classes interaction (see Appendix 1).

[16]See Appendix 1.

TABLE 6A

SUMMARY OF RESULTS OF ANALYSES OF VARIANCE: METHODS, CLASSES, AND CHECK FLIGHTS

(Direction of Mean Differences Between Experimental and Control Groups and Significance of F-Tests)

Photographic Records for Maneuver 7, Straight Climb, and Maneuver 14, Left 180° Turn, 45° Bank

	Man. 7				Man. 14					
	Airspeed Variation	Altitude Difference (Entry)	Altitude Variation (Turn)	Altitude Difference (Recov.)	Slips (Entry)	Slips (Turn)	Slips (Recov.)	Skids (Entry)	Skids (Turn)	Skids (Recov.)
Methods group showing better performance*	C	E	C	E	E	E	E	C	E	=
Variance										
Methods	.05-.01	n.s.**	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Classes	n.s.	n.s.	n.s.	<.01	<.01	<.01	.05-.01	.05-.01	<.01	<.01
Check flights	<.01	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Methods x Classes	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	.05-.01
Methods x Check flights	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Classes x Check flights	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

*When the experimental group shows better performance it is indicated by E; when the control group shows better performance it is indicated by C; when there is no difference it is indicated by =. Of course, this only shows the direction and does not necessarily mean that the difference is significant.

**Not significant at 5% level of confidence, (nor at 10% level of confidence for methods variance).

TABLE 6B

SUMMARY OF RESULTS OF ANALYSES OF VARIANCE: METHODS, CLASSES, AND CHECK FLIGHTS

(Direction of Mean Differences Between Experimental and Control Groups and Significance of F-Tests)

Photographic Records for Maneuver 15, Right 180° Turn, 45° Bank

	Man. 15							
	Altitude Difference (Entry)	Altitude Variation (Turn)	Altitude Difference (Recov.)	Slips (Entry)	Slips (Turn)	Slips (Recov.)	Skids (Entry)	Skids (Recov.)
Methods group showing better performance*	C	C	C	E	C	C	E	E
Variance								
Methods	n.s.**	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Classes	$<.01$	n.s.	n.s.	$<.01$	$<.01$	$<.01$	n.s.	$<.01$
Check flights	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Methods x Classes	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Methods x Check flights	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Classes x Check flights	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

*When the experimental group shows better performance it is indicated by E; when the control group shows better performance it is indicated by C; when there is no difference it is indicated by =. Of course, this only shows the direction and does not necessarily mean that the difference is significant.

**Not significant at 5% level of confidence, (nor at 10% level of confidence for methods variance).

TABLE 6C

SUMMARY OF RESULTS OF ANALYSES OF VARIANCE: METHODS, CLASSES, AND CHECK FLIGHTS

(Direction of Mean Differences Between Experimental and Control Groups and Significance of F-Tests)

Instructor Grades for Maneuver 7, Straight Climb; Maneuver 14, L. 180° Turn, 45° Bank; Maneuver 15, R. 180° Turn, 45° Bank; and Over-all Grades

	Man. 7	Man. 14	Man. 15	Over-all Grade
Methods group showing better performance*	C	E	E	C
Variance				
Methods	n.s.**	n.s.	n.s.	n.s.
Classes	n.s.	$<.01$	$<.01$	n.s.
Pre-check flights	$<.01$	$<.01$	$<.01$	$<.01$
Methods x Classes	n.s.	.05-.01	$<.01$	$<.01$
Methods x Pre-check flights	n.s.	n.s.	n.s.	n.s.
Classes x Pre-check flights	n.s.	n.s.	n.s.	n.s.

*When the experimental group shows better performance it is indicated by E; when the control group shows better performance it is indicated by C; when there is no difference it is indicated by =. Of course, this only shows the direction and does not necessarily mean that the difference is significant.

**Not significant at 5% level of confidence, (nor at 10% level of confidence for methods variance).

each of the instructor criterion measures there are significant check flight variances. This, in view of the progressive increase of the scores from check flight to check flight, is probably a manifestation of learning, and is in accord with the findings in the methods, instructors, and check flights analysis.

Methods x Classes. Significant methods x classes interactions were found for one of the photographic criterion measures, Skids on recovery, on Maneuver 14. Significant methods x classes interactions were found for three of the four instructor criterion measures. These significant methods x classes interactions could result from differences between instructors, as has already been indicated.[17]

Methods x Check Flights and Classes x Check Flights. None of the methods x check flights and classes x check flights was found to show significant variances. This would indicate that the increments in pilot proficiency shown by the student pilots, and reflected in their criterion scores could not be shown to be a function of the type of training received by the subjects or of the class in which they flew.

RESULTS: INSPECTOR CRITERIA

Figures 3, 4, 5, 6, 7, and 8 present the group learning curves for the two inspectors (two classes) on the criterion measures, over-all grades, maneuver grades (for the Straight Climb, the Left and Right Medium Turns), and the scores computed from the inspectors' observations on the Ohio State Flight Inventory for these maneuvers. As has been mentioned before, there is reason to suspect the "reliability" and "validity" of the grades assigned by the inspector of the third class.[18] For this reason more rigorous analysis was not carried out.

For the second class, there appears to be a distinct tendency for the control group to show better performance on all the grades assigned by the inspector.[19]

For the grades assigned by the inspector of the third class, however, the results are not clear cut.[20] Neither group shows any consistent superiority over the other group. The means of the two groups are fairly close for each check flight and the rank order of the means of the two groups tends to change from check flight to check flight and from criterion to

---

[17] See Appendix 1.

[18] See Footnote 7.

[19] See Figures 3, 4, and 5.

[20] See Figures 6, 7, and 8.

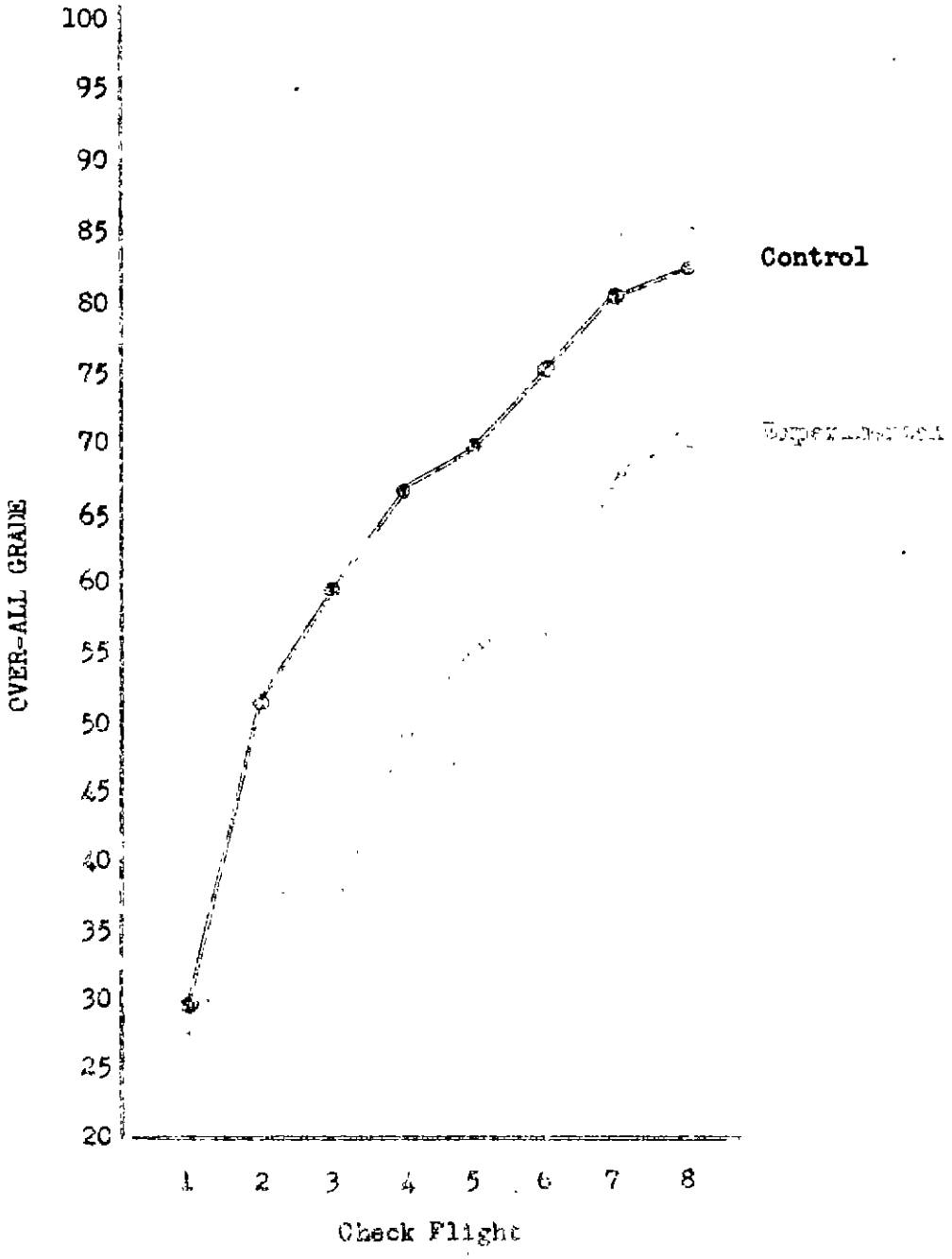

FIGURE 3

OVER-ALL GRADE BY CHECK FLIGHT
(Inspector of 2nd Flight Class)

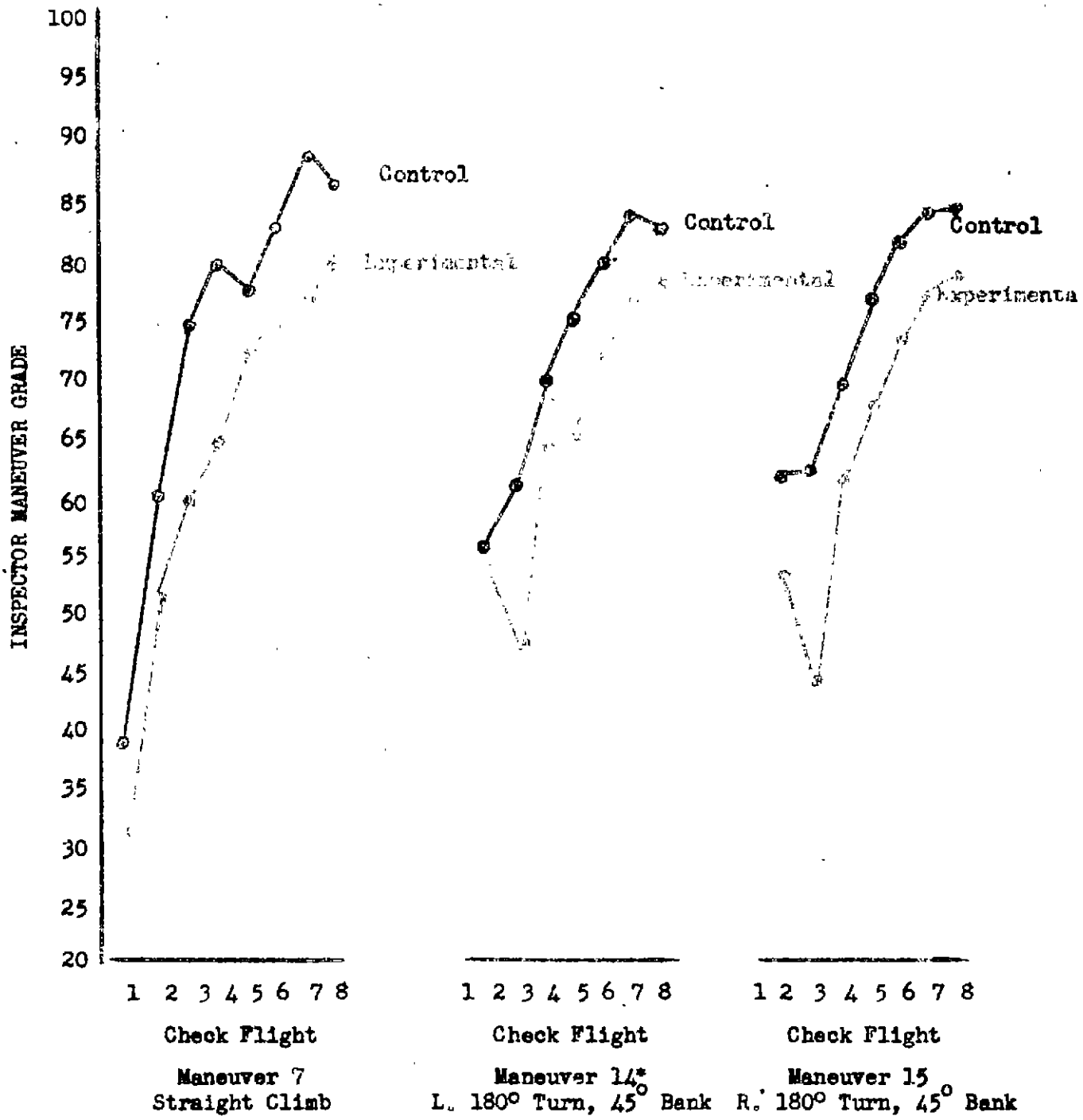

FIGURE 4

MANEUVER GRADES BY CHECK FLIGHT
(Inspector of 2nd Flight Class)

*Maneuver 14 was not included in the first check flight.

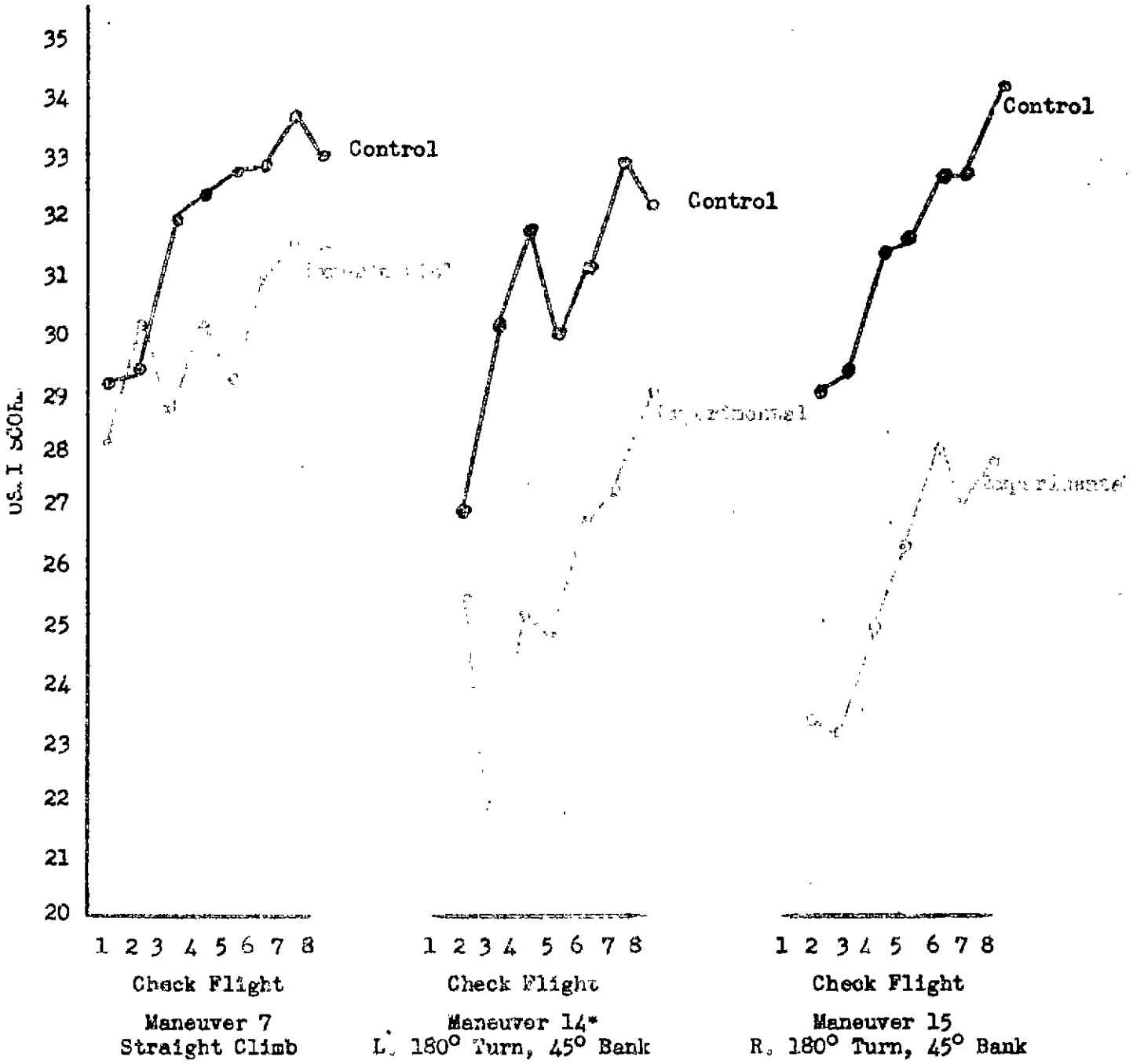

FIGURE 5

OSFI SCORE BY CHECK FLIGHT
(Inspector of 2nd Flight Class)

*Maneuver 14 was not included in the first check flight.

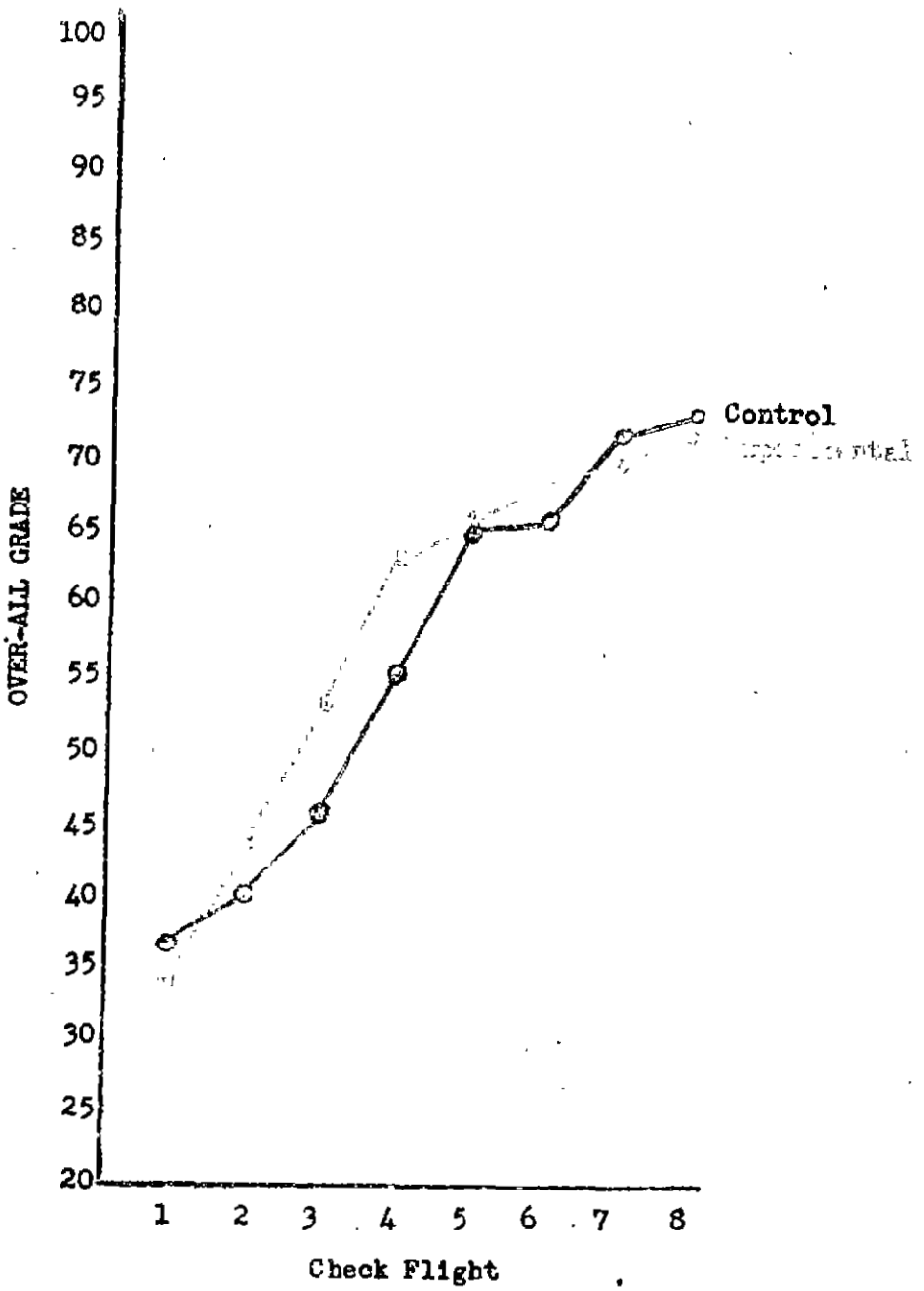

FIGURE 6

OVER-ALL GRADES BY CHECK FLIGHT
(Inspector of 3rd Flight Class)

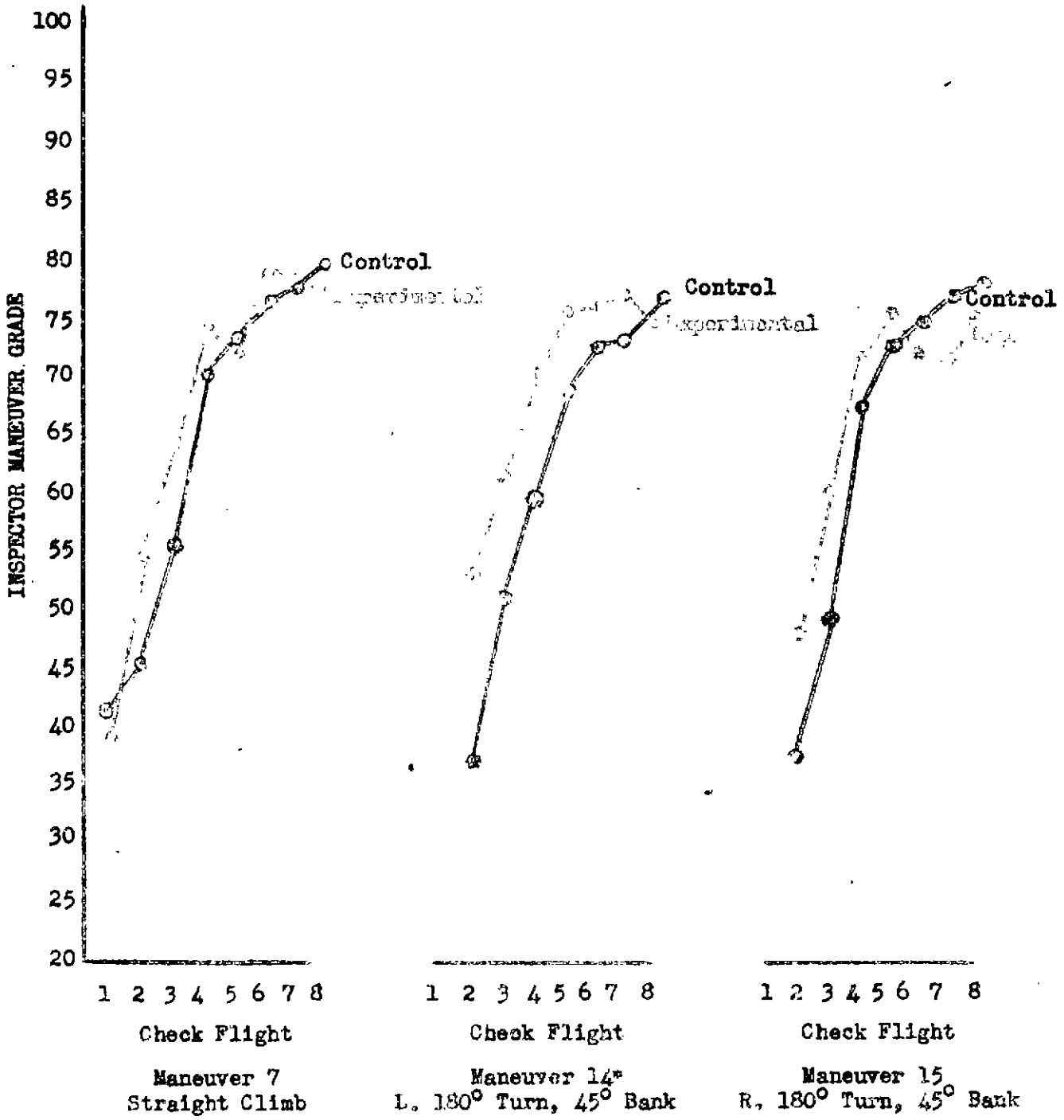

FIGURE 7

MANEUVER GRADES BY CHECK FLIGHT
(Inspector of 3rd Flight Class)

*Maneuver 14 was not included in the first check flight.

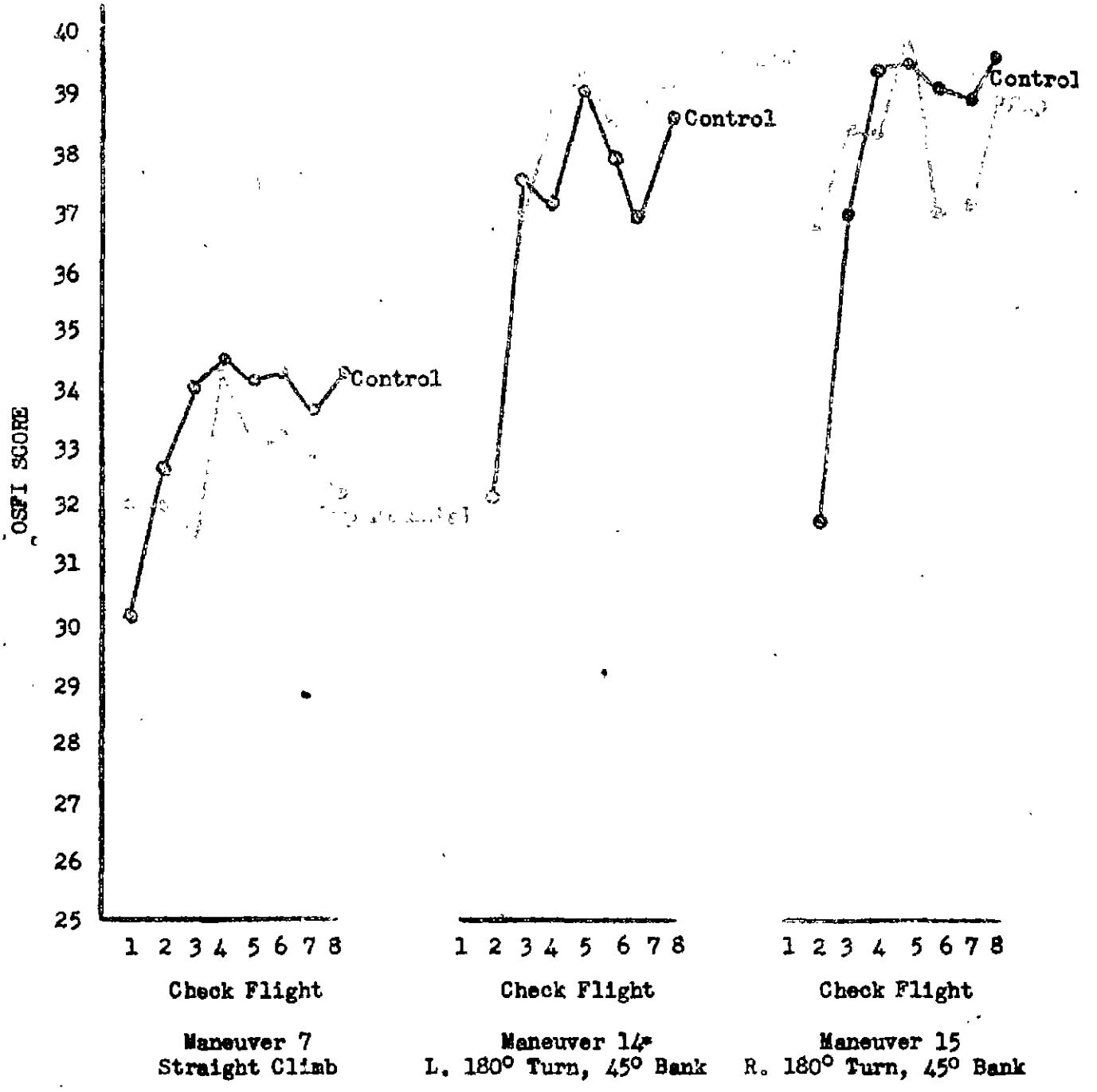

FIGURE 8

OSFI SCORES BY CHECK FLIGHT
(Inspector of 3rd Flight Class)

*Maneuver 14 was not included in the first check flight.

criterion. The control group seems to show better performance than the experimental group on the 8th check flight for the over-all grades, each of the maneuver grades, and each of the Ohio State Flight Inventory scores except for Maneuver 14, Left Medium Turn

Although there is little difference between the experimental and control groups in the third class, there is a consistent difference in the second class in favor of the control group. This may be taken as further evidence of the fact that no marked differences in efficiency between the two procedures exists, and supports the suggestion that if there is a true difference between methods it is apparently in favor of the control group.

DISCUSSION

In spite of the limitations in the design of this experiment, the results indicate quite forcibly that subjects trained by the "experimental" and "control" procedures did not differ in important nor significant degree as far as any of the varied criteria of flight proficiency were concerned. In terms of most of the criterion variables the control subjects were, in fact, slightly superior to the experimental subjects, although the differences for the most part were not statistically significant. The few statistically significant "methods" differences which were found were all in favor of the control group. The possibility exists that these significant differences reflect an instructors x methods interaction, although it seems more plausible to assume that a methods difference was involved. In any case the results of this experiment do not indicate superiority of flight performance resulting from use of the experimental training procedure. These conclusions are also supported by an examination of the analysis made of the flight test grades assigned by the inspectors.

There are, however, several cautions which should be indicated in the interpretation of results. The experiment was limited only to the first 35 hours of training, and it is possible that if criterion measures based on additional flight training were obtained, the results might have been different. Furthermore, it is possible that the experimental training might produce better instrument pilots. These cautions, however, need not be taken too seriously as far as training of private pilots is concerned since most private pilots do not require much more than 35 hours of flight training in a controlled course (i.e., not more than 40 or 45 hours), and since not a great number of private pilots go on to the instrument rating. Another serious limitation is, of course, that no definitive information is available indicating the degree to which the instructors absorbed their indoctrination, or the degree to which the experimental and control instruction actually differed. Furthermore, the precision of the experiment is impaired to the degree that the assumptions of "no instructor differences" and "no class differences" are not gratified

Although the results of this experiment, regarding the relative efficacy of two types of instructional procedure, are not altogether unequivocal, nevertheless, the analyses of these data yielded a number of implications bearing upon future experimentation in this field. For example, since there were significant variances associated with both the instructors

[illegible] the criterion
tions in which [illegible] engage-
ment so that [illegible] simultaneously.

It is also [illegible] on the assumption of [illegible] tor variance was [illegible] on the instructors' grade, than [illegible] judgment (i.e., the [illegible] suggest that in analyses of criteria based upon instructors' grades, such instructor variance as is found is due primarily to differences in grading standards and procedures [illegible] standards and procedures, i.e., differences among instructors in weather, ability. If this were true it would indicate that criteria should be placed on objective or rigorously derived measures [illegible] or upon the qualitative judgment of a single individual, [illegible] not engaged in instruction of subjects in the experiment in question.

This suggestion must be qualified, however, by a number of considerations. Criterion data based upon the instructors' grade were taken at each of the eight check flights, i.e., at the fifth, tenth, fifteenth, twentieth, twenty-fifth, thirtieth, thirty-fourth, and thirty-fifth hour of flight training. Criterion data based upon the photographic records were taken only at the check flights occurring at ten, twenty, thirty-four, and thirty-five hours, and the data from the second check flight (at ten hours) were very sparse. Because of this fact, as has been noted previously, in the analysis by "methods, instructors, and check flights" photographic data from the second check flight were in general excluded. The fact that a greater proportion of significant variances were evident for criterion measures based on instructors' grades might, therefore, be due to the fact that these measures more adequately sampled performance at intervals early in the course.

Similarly, because of this possibility, the fact that among the photographic criteria there were few instances of significant variances attributable to check flights must be interpreted guardedly. This might be considered to reflect upon the discriminatory power of criterion measures as "atomistic" in nature (i.e., having reference to specific elements of performance) as are these photographic criteria. This would follow since apparently increments in flight proficiency were not registered by these measures. However, since the photographic measures were for the most part taken in the latter part of the course (during the last 15 hours of the 35-hour course), meaningful comparisons cannot be made between the incidence of significant check flight variances occurring among criterion measures based respectively upon the instructors' grades and upon the photographic records. The fact that at least in the last 15 hours of the 35 hour course increments in flight proficiency were not registered in terms of the photographic criteria employed in this study is, however, of some significance.

---

[21]Although data from the second check flight were too sparse to permit a breakdown by instructors, data from this check flight were included in the "methods, classes, and check flights" analyses. However, the paucity of data from this check flight, and the lack of precision induced by the limitations of the experimental design, renders particularly suspect conclusions regarding performance early in the course.

APPENDIX 2

APPENDIX 1

Implications of Experimental Design. It is important to examine some of the implications of the experimental design in order to evaluate the analysis which was eventually carried out. The design is presented schematically in Figure 9.

	Class 2	Class 3
Instr 1	X_{1C2}	X_{1E3}
Instr. 5	X_{5E2}	X_{5C3}

FIGURE 9

For purposes of exposition, the check flight variable has been dropped and only two instructors are considered, Instructor 1 and Instructor 5.

X_{IMC} denotes the mean of the criterion scores for students having a particular instructor subjected to one of the methods of training and being members of the same class. For example, X_{1E3} denotes the mean criterion score of students who were taught by Instructor 1, who were given the experimental training and who were in the third class.

This schematic representation shows the basic characteristics of the experimental design. It may be seen that for Instructor 1 the control students are in the second class, and for Instructor 5, the control students are in the third class. The experimental students for Instructor 1 are in the third class and are in the second class for Instructor 5.

This type of experimental design presents certain difficulties. It may be shown, for example, that a difference between methods could be a reflection of an instructors x classes interaction, and vice versa. The possible confusions that may occur are presented by means of algebraic transformation of inequalities following:

1. A methods difference could be a reflection of an instructors x classes interaction and vice versa

$X_{1E3} + X_{5E2} \neq X_{1C2} + X_{5C3}$ (Methods difference)

$X_{1E3} + X_{1C2} \neq X_{5C3} + X_{5E2}$

$X_{5E2} + X_{1C2} \neq X_{5C3} + X_{1E3}$ (Instructors x Classes interaction)

2. A classes difference could be a reflection of a methods x instructors interaction, and vice versa.

X_{1E3} [illegible]

X_{1E3} [illegible] (... interaction)

X_{1E3} [illegible]

3. An instructors [illegible] methods x classes interaction, [illegible]

$X_{1E3} + X_{1C?} \neq X_{5E2} + X_{5C3}$ (Instructors difference)

$X_{1E3} - X_{5E2} \neq X_{5C3} - X_{1C2}$

$X_{1E3} - X_{5C3} \neq X_{5E2} - X_{1C2}$ (Methods x Classes interaction)

It should be noted that in the algebraic traneformations, hypothetical obtained values of the mean criterion measures are represented. These transformations do not necessarily imply that tno differences and interactions are significant.