

ANALYSIS OF RATINGS IN THE PRELIMINARY
PHASE OF THE C.A.A. TRAINING PROGRAM

by

H. M. Johnson

and

Mary L. Boots

A report on research conducted at Tulane University,
New Orleans, Louisiana, by means of a grant-in-aid from
the National Research Council Committee on Selection and
Training of Aircraft Pilots from funds provided by the
Civil Aeronautics Administration.

October 1943

CIVIL AERONAUTICS ADMINISTRATION

Division of Research

Report No. 21

Washington, D. C.

National Research Council
Committee on Selection and Training of Aircraft Pilots
Executive Subcommittee

M. S. Viteles, Chairman

C. W. Bray	J. C. Flanagan
D. R. Brimhall	H. M. Johnson
D. J. Brimm	W. R. Miles
L. A. Carmichael	G. R. Wendt
J. W. Dunlap	

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

September 22, 1943

Dr. Dean R. Brimhall
Director of Research
Civil Aeronautics Administration
Washington, D. C.

Dear Dr. Brimhall:

Attached is a report entitled Analysis of Ratings in the Preliminary Phase of the C.A.A. Training Program by H. M. Johnson and Mary L. Boots. The report is submitted by the Committee on Selection and Training of Aircraft Pilots with the recommendation that it be included in the series of technical reports issued by the Division of Research, Civil Aeronautics Administration.

It will be noted that this report embodies the results of one of the earliest investigations of the efficacy of the methods of rating flight performance in use in the early phases of the Civilian Pilot Training Program. The report is of current interest because it includes considerable data concerning the reliability of instructor and inspector ratings and of the relationships between them.

Cordially yours,



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

MSV:rm

20-11-43

EDITORIAL FOREWORD

Research on the selection and training of pilots involves a careful analysis of methods used in assessing the flight proficiency of student pilots. For this reason, studies of the reliability and validity of ratings given by flight instructors and inspectors were included among the first investigations sponsored by the Committee on Selection and Training of Aircraft Pilots.

The study by H. M. Johnson and Mary L. Boots represents one of the earliest investigations in this field. It is of particular interest because it involves an evaluation of methods actually employed by those responsible for giving instruction and for deciding whether a student pilot should or should not be granted a license to fly a plane. The findings of the investigation were employed in directing the efforts of the Committee and of the Civil Aeronautics Administration in developing criteria of achievement in learning to fly more suitable than those employed in the early phases of the Civilian Pilot Training Program. While some of the findings may appear to be "out of date," because of developments subsequent to 1940, the study and its findings are nevertheless of interest in relation to the problems experienced and the methods currently employed in the assessment of flight proficiency.

CONTENTS

	Page
EDITORIAL FOREWORD	v
SUMMARY.	ix
INTRODUCTION	1
SUBJECTS	1
THE RATINGS.	1
RESULTS.	4
Intercorrelations between stages in training . . .	5
Correlations between inspectors' ratings and instructors' mean ratings.	6
Correlations between inspectors' ratings and instructors' mean ratings on maneuvers for last two hours only.	10
Maneuver intercorrelations	11
Correlations for students with previous training .	17
Time in stages of training	17
"REMARKS" BY INSTRUCTORS	19

SUMMARY

An analysis was made of the ratings given by instructors and inspectors to students participating in the experimental program for Civilian Pilot Training conducted by the Civil Aeronautics Authority in the spring of 1939. The major portion of the analysis covered records of 178 students without previous flight training. A supplementary analysis was made of the records of 110 students who had had previous flight training.

Ratings by instructors were made on a number of maneuvers during every lesson. All ratings under certain general headings were averaged for each stage of training. Ratings by inspectors were made on a percentage basis. The actual range of ratings for students with no previous training ranged from 70 to 95.

Average ratings for the majority of students tended to progress from stage to stage, although there were numerous exceptions to this tendency.

Positive and significant intercorrelations were found between ratings at various stages. These correlations ranged from .88 for two adjacent stages to .44 for stages removed by two intervals.

Correlations between inspectors' final general ratings and instructors' mean ratings on given maneuvers for the entire course were low, although a few were significant. The higher correlations were found with maneuvers which had been given the largest number of grades and so might be expected to have higher reliability. Doubt is expressed as to the practicality of increasing reliability by increasing the number of ratings since under the current system the number of grades given to one student throughout the course ranges from approximately 600 to over 1500. A suggestion is made that only ratings of "passing" or "not passing" be required.

Correlations between inspectors' final general ratings and instructors' mean ratings on maneuvers for the final two hours of flight instruction tended to be somewhat higher, although the highest was only .33.

Intercorrelations between instructors' mean ratings for given maneuvers ranged from .29 to .93. (Here again, the higher correlations tended to be between maneuvers more frequently rated.)

Correlations on a group of 110 students with previous training were computed separately. These correlations tended not to be greatly different from those for students without previous training.

An analysis of remarks made by instructors on the rating scales revealed many valuable and enlightening suggestions for the use of the student or of another instructor. However, very frequently the remarks proved to be perfunctory or to contradict the ratings. It is suggested that both the system of rating and the requirement for explanatory remarks should be not too intricate or too involved.

ANALYSIS OF RATINGS IN THE PRELIMINARY

PHASE OF THE C.A.A. TRAINING PROGRAM

INTRODUCTION

In the spring of 1939, the Civil Aeronautics Authority conducted an experimental program for civilian pilot training in thirteen different colleges, with 330 students participating. In November 1939 it was decided to make an analysis of the ratings given these students by their instructors and by the aeronautical inspectors. Such a study, it was felt, should prove or disprove statistically the efficacy of the current rating system; it should expose discrepancies, if they existed, between instructors and inspectors in the use of the system; it should demonstrate whether this system provided the correct standard for judging the proficiency or advancement of students; it should reveal the relative importance of different maneuvers, or the emphasis put upon them by instructors; and thereby, should be an aid to instructors and to those responsible for planning future courses in civilian pilot training.¹

SUBJECTS

Records were analyzed of students in the twelve institutions listed in Table I.² Of these 300 who started the course, twelve were washouts, but of these twelve, as far as available records show, only two actually finished the course and were disapproved by the aeronautical inspector.

It was then decided to eliminate from this total of 288 all those who had had previous training of any kind, either in dual or solo flying, or in glider flights. Records³ showed that 110 students had had some previous experience, and their elimination left a total of 178 without previous training to be treated statistically in one group. The records of students with previous training received a separate analysis.

THE RATINGS

The flight training time was divided into Stage A, Dual Instruction, minimum 8 hours; Stage B, Primary Solo, minimum 1 hour dual and 5 hours solo; and Stage C, Advanced Solo, minimum 8 hours dual and 13 hours solo. The ra-

¹Student records were procured through the courtesy of Dr. Dean R. Brimhall, Director of Research, Civil Aeronautics Administration. Particular methods of study were determined with the help of Dr. V. A. C. Gannon, Dr. Percy W. Cobb, and Dr. Raymond Fraumeni.

²The records of 30 students at the University of Washington were incomplete and have been eliminated from consideration.

³Obtained from V. A. C. Gannon.

tings obtained varied somewhat from stage to stage, as indicated in Table II. (In Stage C, there were two rating sheets for grades. The corresponding sub-stages are indicated as C-1 and C-2.⁴)

TABLE I
STUDENT RECORDS ANALYZED FROM THE EXPERIMENTAL PHASE
OF THE VOCATIONAL FLIGHT TRAINING PROGRAM

<u>Colleges</u>	<u>Number Starting Course</u>	<u>Considered for Correlation</u>	<u>Washed out by Instructor</u>	<u>Had Previous Training</u>
University of Alabama	30	24	0	6
Georgia School of Technology	30	21	2	7
University of Kansas	20	13	0	7
Massachusetts Inst. of Technology	20	13	2	5
University of Michigan	20	9	1	10
University of Minnesota	20	10	1	9
New York University	30	17	4	9
North Carolina State College	20	9	0	11
North Texas Agricultural College	30	18	1	11
Pomona Junior College	15	13	1	1
Purdue University	50	18	0	32
San Jose State College	15	13	0	2
	300	178	12	110

The flight instructors graded each of the students on a regulation five-point rating scale, the Civil Aeronautics Authority Flight Record (see exhibits 1 and 2). The ratings were made according to the directions printed on the rating forms (the Flight Records) as follows: 1-good, 2-above average, 3-average, 4-below average, and 5-poor. One of these forms (Flight Records) was available for each stage of flight training. The students were rated in each period of instruction on the sub-headings given in Table II under Stages A, B, and C. (An "x" indicates which maneuvers were rated in a given stage.) It was decided for the purposes of this study to average, without weighting, all the grades given under the general headings, identified by Roman numerals, divided however into Stages A, B, and C-1 and C-2. Stage A was also divided into four periods approximately equal in time in order to correlate the ratings in the very early hours of dual training with the ratings in the later periods of dual training. A form was made up to show the average grade given to each student on each maneuver in each stage of instruction, the average grade for each stage (all maneuvers), and the general average of all maneuvers for the whole course.

⁴Stage C-2 in the early phase of the C.A.A. training program corresponds roughly to Stage D in the later C.P.T. course.

STUDENT

NAME _____
 AIR _____
 INSTRUCTOR _____
 INSTRUCTOR _____
 RATING HELD _____
 SOLO HOURS _____

CIVIL AERONAUTICS AUTHORITY

FLIGHT RECORD

GRADE A - REAL INSTRUCTION
 MINIMUM 8 HOURS DUAL

STUDENT

AIRCRAFT CERTIFICATE _____
 MAKE & MODEL _____
 TYPE _____
 CLASSIFICATION _____
 (NEW/OLD & ENGINE) _____

FLIGHT NUMBER	I			II			III			IV			V			VI			VII			VIII			IX			DATE	YEAR	DAY	INSTRUCTOR'S ATTITUDE AND PROGRESS	INSTRUCTOR'S VERIFICATION
	STANDING	STARTING	ENDING	STARTING	ENDING	ENDING	STARTING	ENDING	ENDING	STARTING	ENDING	ENDING	STARTING	ENDING	ENDING	STARTING	ENDING	ENDING	STARTING	ENDING	ENDING	STARTING	ENDING	ENDING								
1																																
2																																
3																																
4																																
5																																
6																																
7																																
8																																
9																																
10																																
11																																
12																																
13																																
14																																
15																																
16																																
17																																
18																																
19																																
20																																

STUDENTS SHOULD BE PERMITTED TO SOLO AT ANY TIME AFTER 8 HOURS WHEN IN THE OPINION OF THEIR INSTRUCTORS THEY ARE QUALIFIED. THE ORDER OF TEACHING AND TIME SPENT ON THE ABOVE MANEUVERS IS LEFT TO THE INSTRUCTOR'S JUDGMENT TO SUIT HIS CONDITIONS AND PERSONNEL. INSTRUCTION PERIODS ARE 30 MINUTES EACH. APPROACHES TO BE 90° AND 180° U-TYPE WITHOUT POWER.

PROPER ENTRIES INCLUDING GRADE ON EACH MANEUVER PRACTICED WILL BE MADE AFTER EACH FLIGHT BY THE INSTRUCTOR, USING FOLLOWING GRADING SCALE: 1 - GOOD, 2 - ABOVE AVERAGE, 3 - AVERAGE, 4 - BELOW AVERAGE, 5 - POOR. RECORDS WILL BE KEPT IN QUADRICULATE: 1 TO C.A.A., 1 TO COLLEGE, 1 TO INSTRUCTOR, 1 TO STUDENT LOG.

I HEREBY CERTIFY THAT THE INSTRUCTION AND SOLO TIME HAS BEEN GIVEN AS INDICATED ABOVE.

I HEREBY CERTIFY THAT THE INSTRUCTION AND SOLO TIME HAS BEEN RECEIVED BY ME AS INDICATED ABOVE.

INSTRUCTOR

STUDENT

[illegible]

12

I HEREBY CERTIFY THAT THE INSTRUCTION AND SOLO TIME HAS BEEN RECEIVED BY ME AS INDICATED ABC

I HEREBY CERTIFY THAT THE IN-
STRUCTION AND SOLO TIME HAS BEEN
GIVEN AS INDICATED ABOVE.

1. IN STAGE C, LENGTH OF PERIOD SHOULD RANGE FROM 30 MINUTES TO ONE HOUR AT DISCRETION OF INSTRUCTOR AFTER JUDGING STUDENT AND HIS TENDENCY TO FIRE. ADDITIONAL TIME TO THAT OUTLINED IN THE LEFT HAND MARGIN SHOULD BE GIVEN ON ANY MANEUVER THE INSTRUCTOR DEEMS NECESSARY. ABOVE IS A MINIMUM OUTLINE ONLY. STIMULATED FORCED LANDINGS SHOULD BE GIVEN ON ALL DUAL PERIODS.

INSTRUCTOR

EXHIBIT 2

TABLE II

MANEUVERS RATED DURING EACH STAGE⁵

<u>Maneuver</u>		<u>Stage A</u>	<u>Stage B</u>	<u>Stage C-1</u>	<u>Stage C-2</u>
I.	PRELIMINARY GROUND INSTRUCTION				
	Starting engine	x	x	x	x
	Warming up engine	x	x	x	x
	Stopping engine	x	x	x	x
	Warnings	x			
II.	TAXIING				
	Into wind	x	x	x	x
	Cross wind	x	x	x	x
	Down wind	x	x	x	x
	Gusty air	x			
III.	TAKE-OFFS				
	Into wind	x	x	x	x
	Cross wind	x	x	x	x
IV.	AIRWORK				
	Straight and level	x	x	x	x
	Gentle glides and climbs	x	x	x	x
	Gentle turns	x	x	x	x
	70° turns	x	x	x	x
	Spirals and approaches	x	x	x	x
V.	LANDINGS				
	Into wind	x	x	x	x
	Cross wind	x	x	x	x
VI.	STALLS AND SPINS	x			
VII.	EMERGENCIES				
	On take-off	x	x	x	x
	90° to wind	x	x	x	x
	180° to wind	x	x	x	x
VIII.	SOLD (3 landings)	x			
X-XII.	PRECISION LANDINGS, EIGHTS, TURNS WITH 70° BANK				
	180° landings			x	x
	30° eights			x	x
	360° landings			x	x
	70° power turns			x	x
	720° landings			x	x
	70° eights			x	x
XIII.	EMERGENCY MANEUVERS				
	Stalls			x	x
	Spins			x	x
	Slips			x	x
XIV-XV.	CROSS-COUNTRY				
	Power approaches				x
	Dragging field				x
	Power landings				x
	Cross-country flight				x
XVI.	FRONT SEAT INDOCTRINATION AND PRIVATE TEST PRACTICE				x
XVII.	FLIGHT EXAMINATION BY INSTRUCTOR				x

⁵An "x" indicates which maneuvers were rated in a given stage.

The inspectors' ratings used throughout the study are their final general ratings, made on a percentage basis. No ratings were available that were given by inspectors on the performance of individual maneuvers in the examination for certification.

Since correlations were the chief method of attack, it must be noted that the range of ratings is very small. An instructor almost never gave the rating 5 and very infrequently gave the rating 4. The scale then in most cases became a 1, 2, 3 rating scale. The aeronautical inspectors' range was 70-100; their "passing" mark was 70; only one student received 95, the highest grade awarded in this group of 178 students who had had no previous training.

It must also be noted that the ratings given in twelve different schools by forty different instructors have been combined in setting up the frequencies for these correlations. At one school many 1's may be given; at another school, the small number of 1's and the very large number of 2's seem to indicate different concepts of the rating scale of 1, 2, 3, 4, 5. On the other hand, it seems entirely valid to combine the 178 students from the 12 different schools because the regulations and instructions for rating were given uniformly to all instructors by one central authority.⁶ This is not the case in many statistical comparisons of college students under instructors who have individual ideas of rating. It is true that it is possible to reduce the degree of correlation by mingling uncorrelated pairs with correlated pairs, but it is also true that two elements uncorrelated in each of two records, will show some degree of correlation in a mingling of the two records, unless the mean values are identical in the two records. The frequencies under consideration are too small to afford valid testing of these two theorems.⁷

RESULTS

In studying the records the question arose as to whether the instructor started always with the lower ratings and progressed to the higher ratings during the course, or whether, as maneuvers became more advanced, lower ratings were given. A simple count shows that of these 178 students, 12 of them scattered among six institutions had lower averages on the last quarter of Stage A than on the first quarter. The inspectors' ratings for these 12 ranged from 70 to 86. Six students in four different colleges had the same average for A-4 as for A-1. Their ratings by the inspectors ranged from 74 to 93. This leaves a total of 160 who had progressed to higher ratings at the end of Stage A. In a comparison of the ratings on Stage C and Stage A, 24 students in seven different colleges had lower averages for Stage C than for Stage A. The inspectors' ratings for these 24 students ranged from 70 to 86. There were 15 students in six different places of training who had the same average rating for Stage C

⁶ Editor's Note. It must be noted that the instructions may not have been adequate to assure uniformity of interpretation.

⁷ Editor's Note. It has been suggested that all assumptions regarding scale units, and similarity of interpretation of the various points along the scale by the different instructors could be avoided by comparing the inspectors' ratings of the students placed by each instructor in the top and bottom quarter of his group.

as for Stage A, and their grades from the inspectors ranged from 75 to 95. Thus, a total of 139 had higher averages for the last stage than for the first stage. A count was made for the maneuvers VII to XVII of Stage C. The result showed that 60 students in nine different colleges had a lower average for these late maneuvers than for Stage A and the range of inspectors' ratings for them was 70 to 95. Eighteen students in nine different schools had the same average as for Stage A, with a range of 70 to 87 in inspectors' ratings. A total of 100 students then had higher averages for C-VII to XVII than for Stage A.

One other count was made. Comments have been heard that instructors commonly give grades on practice test flights lower than their ratings on maneuvers, possibly to impress the inspectors. This count showed that nine students had not been rated on test practice, and of the remaining 169, there were 104 who did receive lower ratings for test practice than their general mean on all maneuvers for the whole course. Eighteen had the same average and 47 had a test-practice rating that was higher than their general mean.

Another question that came up was whether the instructors on this program had been re-rated. An examination of their files shows that out of the 40 instructors who signed these records, 27 were re-rated and held instructor's certificates the following year. Certificates had expired for 13, one each from Massachusetts Institute of Technology, Michigan, Minnesota, North Carolina, Pomona and Texas, two each from Georgia and Purdue, and three from Georgia.

Intercorrelations between stages in training. In order to ascertain to what extent instructors' ratings on performance in the first few hours of training would predict ratings of performance in the later hours of dual training, and what prediction could be made of ratings of performance in the solo stages of training by ratings during the dual stage, the nine intercorrelations listed in Table III were computed. It is seen that consecutive stages yield the highest correlation. There is a closer relationship between Stages A and B, and also between Stages B and C than between Stage A and Stage C. In the correlation of ratings made in Stage A with Stage C ratings, the value of r is .54, and r^2 is .29, which means that only 29% of the variation of ratings in Stage C can be "explained" in terms of the Stage A ratings. The other 71% must be explained by other factors. It must be recognized, of course, that the reliability of ratings in a single stage is far from perfect.

The same situation holds true for the four periods of Stage A. Stage A-1 and Stage A-2, Stage A-2 and Stage A-3, Stage A-3 and Stage A-4 have correlation coefficients of .71, .82 and .78 respectively; Stage A-1 and Stage A-3, Stage A-2 and Stage A-4, each with one intervening stage, have .60 and .64 respectively; while the correlation of Stage A-1 and Stage A-4, with two intervening stages, has the lowest coefficient, namely .44. All these coefficients, however, were highly significant. It is possible that some students who "catch on" slowly, or who are nervous and tense at the start, may attain the same degree of proficiency at the end of eight hours of dual training as

TABLE III

INTERCORRELATIONS BETWEEN STAGES⁸ IN TRAINING

(Mean of ratings on all maneuvers given student pilots
without previous flight training)

<u>No.</u>	<u>N</u>	<u>r</u>	<u>σ_r</u>	<u>$\frac{r^2}{\sigma_r}$</u>
1. Stage A and Stage B	178	.67	.04	16.44
2. Stage A and Stage C	178	.54	.05	10.26
3. Stage B and Stage C	178	.81	.03	31.19
4. Stage A-1 and Stage A-2	178	.71	.04	19.27
5. Stage A-1 and Stage A-3	178	.60	.05	12.50
6. Stage A-1 and Stage A-4	178	.44	.06	7.16
7. Stage A-2 and Stage A-3	178	.82	.03	32.72
8. Stage A-2 and Stage A-4	178	.64	.04	14.57
9. Stage A-3 and Stage A-4	178	.78	.03	26.97

the students who learn more quickly, or who are not handicapped by tenseness. In general, however, ratings in later stages tend to correlate with ratings in earlier stages.

Correlations between inspectors' ratings and instructors' mean ratings.
The correlations numbered 10 to 25 (see Table IV)¹⁰ were computed to find out what general agreement existed between instructors' ratings and the flight inspectors' ratings. All the coefficients are very low but a few of them are significant. The highest correlation is #13, that between the inspectors' general rating and the instructors' mean rating on Airwork. It is also significant. The only other significant correlations between instructors' ratings and inspectors' ratings, are Landings, Taxiing, and the general mean of all ratings given by the instructors.

⁸Stage A was divided into four periods approximately equal in time.

⁹The ratio, r/σ_r , is a test of the hypothesis that there is no relationship, the true value of r being zero. This ratio is in units of the standard deviation and reference is made to the table of areas under the normal curve. If it is a reasonable criterion of the limit of significance that there is only one chance in 100 that the true value of r is zero, the corresponding value of r/σ_r is 2.576. A value of r/σ_r higher than 2.576 is significant, the probability being small that the true correlation is zero.

¹⁰Intercorrelations are numbered consecutively through Tables III, IV, V, VI, VII, and VIII to facilitate references in the text.

TABLE IV

CORRELATIONS BETWEEN INSPECTORS' RATINGS AND
INSTRUCTORS' MEAN RATINGS

(Mean of ratings on all maneuvers given student pilots
without previous flight training)

Inspectors' rating and: ¹¹	N	r	Cr	$\frac{r}{\sigma_r}$
10. Groundwork (I) mean rating	178	.17	.07	2.37
11. Taxiing (II) mean rating	178	.20	.07	2.83
12. Take-off (III) mean rating	178	.17	.07	2.32
13. Airwork (IV) mean rating	178	.22	.07	3.08
14. Landings (V) mean rating	178	.21	.07	2.93
15. Stalls, Spins (VI) mean rating	171	.00	.08	0.04
16. Emergencies (VII) mean rating	178	.17	.07	2.33
17. Solo (VIII) rating	93	-.05	.10	0.47
18. Precision landings (X-XII) mean rating	178	.14	.07	1.97
19. Emergency maneuvers (XIII) mean rating	178	.13	.07	1.74
20. Cross-Country (XIV-XV) mean rating	162	.09	.08	1.10
21. Test practice (XVI) mean rating	169	.07	.08	0.88
22. Instructors' Flight exam. (XVII) mean rating	156	.13	.08	1.67
23. Instructors' Flight exam. (XVII) (last grade)	156	.11	.08	1.38
24. C-VII to XVII mean rating ¹²	178	.17	.07	2.32
25. Instructors' general mean rating	178	.19	.07	2.60

An interpretation of the results of the correlations between inspectors' ratings and the mean of instructors' ratings on various maneuvers is difficult. In fact, the small range and some departure from a normal distribution make any exact interpretation impossible. If it could be assumed that the higher correlations show greater emphasis put upon those maneuvers by instructors than upon other maneuvers which have correlated lower, in other words, the greater the emphasis the higher the correlation, this reasoning, carried to its logical conclusion, might indicate that not nearly enough

¹¹The Roman numerals identify the maneuver with the listing on the student record forms.

¹²For correlation #24, maneuvers VII to XVII were selected as those which seemed to include the items on the Pilot Flight Test Report graded by the aeronautical inspectors.

TABLE IV-A

CORRELATIONS BETWEEN INSPECTORS' RATINGS AND INSTRUCTORS' MEAN RATINGS ON MANEUVERS IN THE ORDER OF THEIR "r" VALUE¹³

(Mean of ratings on all maneuvers given student pilots without previous flight training)

	<u>r</u>		<u>r</u>
Stalls, Spins (VI)	.00	Take-off (III)	.17*
Solo (VIII)	-.05	Emergencies (VII)	.17*
Test practice (XVI)	.07	Groundwork (I)	.17*
Cross-Country (XIV-XV)	.09	General mean of all ratings by Instr.	.19**
Flight exam. last grade (XVII)	.11	Taxiing (II)	.20**
Emergency maneuvers (XIII)	.13	Landings (V)	.21**
Flight exam. (XVII) mean rating	.13	Airwork (IV)	.22**
Precision landings (X-XII)	.14		
Mean rating on maneuvers VII-XVII in Stage C	.17*		

emphasis was put upon any maneuver. More emphasis need not necessarily mean more total time but it would apparently mean more time spent upon the emphasized maneuver than upon a less emphasized maneuver.¹⁴ Students were given approximately 300 grades on Airwork alone throughout the whole course, but it must be remembered that one rating can be given for a half-minute of performance and also one rating for a half-hour of performance. Sometimes 8 or 10 grades were given on different maneuvers in a 10-minute flight. Not even guess-work could allot the time given to individual maneuvers.

It seems that since time is not a possible element for interpretation, the next consideration is the number of ratings. Inspection of Table IV-A shows that the highest correlations resulted with the maneuvers that had been given the largest number of grades. Those correlations are with Airwork, Landings, Taxiing, Instructors' general mean, Groundwork, Emergencies, Take-off, and the mean of C-VII to XVII. The first four have already been mentioned as significant. If the criterion for the limit of significance is lowered, i.e., if the probability of the true r being zero is set at five times out of 100 (with 1.96 as the corresponding value of $4/\sqrt{r}$) instead of a criterion one time out of 100 (with its corresponding value of r/\sqrt{r} set at 2.576), then the last four correlations men-

¹³ * denotes statistical significance at 2%-5% levels.

** denotes statistical significance at 1% level.

¹⁴ Editor's Note. High correlations between instructors' ratings and inspectors' grades could also result from more emphasis being placed upon certain maneuvers by the inspector, in arriving at his general ratings. This would involve no assumptions regarding amount of time spent on certain maneuvers by the instructors at the various schools.

tioned are also significant, with one other, the correlation of Precision landings, as a border-line case. All of these nine correlations used averages computed from many more ratings than the other seven correlations in this group. Solo had only one rating;¹⁵ Flight exam. (last grade) had of course only one; Flight exam. mean was an average of from two to six grades in cases where more than one rating had been given; Test practice had from four to ten grades; Stalls, Spins, Cross-country, and Emergency maneuvers were given very few ratings, while of these last three, Stalls, Spins, had the fewest.

In correlation #24, maneuvers VII to XVII in Stage C were selected for a grouping because from inspection of the Pilot Flight Test Report they seemed to include the maneuvers judged and graded by the inspector in making up his "Flying Mark." The test-practice grade and the flight exam. grade were included in this grouping because they were ratings given by the instructors on general performance in the last few hours before the inspectors' examination.

The fact that the maneuvers with the larger number of ratings correlate more highly with the inspectors' ratings is in agreement with the usual finding that increasing the number of ratings increases the reliability of the average rating. In the current situation, however, there are practical disadvantages to seeking greater reliability by increasing the number of ratings. With the system of rating in use on this program, the number of grades given to one student throughout his course ranged from approximately 600 to over 1500. From a common-sense point of view it is impossible for an instructor to make that many comparisons with any degree of exactness or even fairness. Just the mechanical setting down of that many grades, without taking the trouble to weigh and balance different factors, is too time-consuming in itself. This view of impracticability and inefficiency has been held by many people, and now in a statistical study, the results of correlations show almost no agreement between this system of rating by instructors and the judgment of the inspectors.¹⁶ In spite of the system, out of the 290 who were presented for examination, only two were disapproved and 288 passed. This means that to a very large degree the judgment of the instructor on the "passing" point only was in agreement with the judgment of the inspector. Why not then require for a rating system only "passing" and "not passing," or "satisfactory" and "unsatisfactory," along with some other term which would show improvement, and thereby assist the instructor? His judgment would be objectified to a great extent; he might still consider each student in a relative position to a whole group who are satisfactory, or another group who are unsatisfactory, but in the main he would not be judging the degree of skill or

¹⁵The number of students who received ratings on Solo is small due to confusion arising from the position of the Solo column on the Stage A, or dual stage, record, instead of on the solo stage record, where it logically belonged. However, this column was eliminated from the records set up for the regular Civilian Pilot Training Program. The Flight exam. by instructor was also eliminated.

¹⁶Editor's Note. It should be noted, however, that the lack of correlation between instructors' ratings and inspectors' grades might result from lack of reliability of the latter, as well as the former.

CORRELATIONS BETWEEN INSPECTORS' RATINGS AND INSTRUCTORS'
MEAN RATINGS ON MANEUVERS FOR LAST TWO HOURS ONLY

(Mean of ratings on all maneuvers given student pilots
without previous flight training)

Inspectors' rating and:	N	r	σ_r	$\frac{r}{\sigma_r}$
26. Groundwork (I)	125	.23	.09	2.75
27. Taxiing (II)	132	.27	.08	3.32
28. Take-off (III)	158	.26	.08	3.40
29. Airwork (IV)	132	.29	.08	3.59
30. Landings (V)	157	.33	.07	4.63
31. Emergencies (VII)	111	.13	.09	1.39
32. Precision landings (X-XII)	150	.11	.08	1.40
33. Emergency maneuvers (XIII)	124	.17	.09	1.91

TABLE V-A

A COMPARISON OF THE CORRELATIONS BETWEEN INSPECTORS' RATINGS AND
INSTRUCTORS' MEAN RATINGS ON MANEUVERS FOR THE LAST TWO HOURS
ONLY WITH CORRESPONDING CORRELATIONS BASED ON THE WHOLE COURSE

(Student pilots without previous training)

	Last 2 hours		Total time	
Inspectors' rating and:	r	T	r	$\frac{r}{\sigma_r}$
Precision landings (X-XII)	.11	1.40	.14	1.97
Emergencies (VII)	.13	1.39	.17	2.33
Emergency maneuvers (XIII)	.17	1.91	.13	1.74
Groundwork (I)	.23	2.75	.17	2.37
Take-off (III)	.26	3.40	.17	2.32
Taxiing (II)	.27	3.32	.20	2.83
Airwork (IV)	.29	3.59	.22	3.08
Landings (V)	.33	4.63	.21	2.93

proficiency of one student in a relative line-up with other individuals.

Correlations between inspectors' ratings and instructors' mean ratings on maneuvers for last two hours only. In the hope of finding closer agreement between instructors' ratings on maneuvers and the general ratings of inspectors, averages were computed for the last two hours only of training. The maneuvers which were rated during the last two hours, together with the correlations between the ratings on these maneuvers and the inspectors' general ratings are listed in Table V-A. In the 3rd and 4th columns are listed the corresponding values for the correlations which used the averages for the whole course.

It is evident that the coefficients for the last two hours are somewhat higher than those for total time with two exceptions, Precision Landings and Emergencies, but these are both insignificant. The highest coefficient, that for Landings, is still only .33, though it is significant. All the ratings considered in these eight correlations were given by the instructors very close in point of time to the general ratings given by the inspectors, yet there is very little agreement.

Maneuver intercorrelations. The 55 intercorrelations listed in Table VI were computed between mean ratings given by the instructors on the various maneuvers. Solo had only one rating for each student and Flight exam. by instructor had only one rating for some students but from two to six ratings for others. However, as mentioned above, these two columns have been omitted from the records in use this year. Test practice had only a few ratings, one for some students, as many as 10 or 12 for others. It can be considered as a rating on general performance which was given some time during the last ten hours. Stalls, Spins, rated in Stage A only, had the fewest grades, very often only three.

In Table VI-A, in which maneuver intercorrelations are listed in the order of the "r" value, it is noticed that Stalls, Spins (Stage A) with Emergency maneuvers (Stage C) has a correlation coefficient of only .44. This is pointed out because "Emergency maneuvers" include under them Stalls, Spins, Slips. This would seem to defeat the proposition that low correlations result because one set of skills has been developed in one maneuver to a degree not attained by other skills in other maneuvers.¹⁷

The last three correlations have coefficients higher than .90: Landings and Airwork, Take-off and Airwork, and Take-off and Landings. They are all infinitely significant. It is likely that proficiency in one member of each pair necessarily has a parallel development of proficiency in the other member of the pair. However, this is not true for the correlation between Emergency maneuvers and Precision landings which is .86. A skill in one may or may not carry with it a corresponding skill in the other, so it might be interpreted as being due to a parallel emphasis on the part of the instructor in training.

For convenience of inspection the coefficients have been arranged in Table VI-B to show the degree of correlation between each maneuver and every other maneuver, and the range of the correlation coefficients for each group.

¹⁷Editor's Note. This evidence is not crucial, since the coefficient of correlation quite probably is attenuated by the lack of reliability of ratings on Stalls, Spins, and Slips, which resulted from the fact that they were rated less often than other maneuvers as pointed out by the authors.

TABLE VI

MANEUVER INTERCORRELATIONS
(Instructors' mean ratings covering all stages)

	N	r	$\frac{r}{\sigma r}$		N	r	$\frac{r}{\sigma r}$		
<u>Take-off (III) and:</u>									
34. Airwork (IV)	178	.92	.01	83.82	<u>Stalls, Spins (VI) and:</u>				
35. Landings (V)	178	.93	.01	84.09	65. Cross-Country (XIV-XV)	157	.37	.07	5.35
36. Stalls, Spins (VI)	171	.51	.06	8.93	66. Test practice (XVI)	163	.29	.07	4.01
37. Emergencies (VII)	178	.80	.03	29.48	67. Flight exam. by Instr. (XVII)	150	.33	.07	4.48
38. Solo (VIII)	93	.69	.05	12.81	<u>Emergencies (VII) and:</u>				
39. Precision landings (X-XII)	178	.78	.03	26.86	68. Solo (VIII)	93	.64	.06	10.54
40. Emergency maneuvers (XIII)	178	.71	.04	19.27	69. Precision landings (X-XII)	178	.81	.03	31.19
41. Cross-Country (XIV-XV)	162	.62	.05	12.92	70. Emergency maneuvers (XIII)	178	.75	.03	22.58
42. Test practice (XVI)	169	.55	.05	10.19	71. Cross-Country (XIV-XV)	162	.69	.04	16.80
43. Flight exam. by Instr. (XVII)	156	.54	.06	9.42	72. Test practice (XVI)	169	.60	.05	12.22
					73. Flight exam. by Instr. (XVII)	156	.60	.05	11.84
<u>Airwork (IV) and:</u>									
44. Landings (V)	178	.91	.01	69.92	<u>Solo (VIII) and:</u>				
45. Stalls, Spins (VI)	171	.57	.05	10.90	74. Precision landings (X-XII)	93	.65	.06	10.87
46. Emergencies (VII)	178	.82	.03	32.76	75. Emergency maneuvers (XIII)	93	.54	.07	7.44
47. Solo (VIII)	93	.68	.06	11.86	76. Cross-Country (XIV-XV)	91	.45	.08	5.31
48. Precision landings (X-XII)	178	.76	.03	24.58	77. Test practice (XVI)	90	.49	.08	6.04
49. Emergency maneuvers (XIII)	178	.68	.04	17.05	78. Flight exam. by Instr. (XVII)	84	.45	.09	5.20
50. Cross-Country (XIV-XV)	162	.57	.05	10.79	<u>Precision landings (X-XII) and:</u>				
51. Test practice (XVI)	169	.50	.06	8.66	79. Emergency maneuvers (XIII)	178	.86	.02	45.26
52. Flight exam. by Instr. (XVII)	156	.50	.06	8.35	80. Cross-Country (XIV-XV)	162	.77	.03	24.16
					81. Test practice (XVI)	169	.74	.03	21.85
<u>Landings (V) and:</u>									
53. Stalls, Spins (VI)	171	.55	.05	10.38	82. Flight exam. by Instr. (XVII)	156	.68	.04	15.91
54. Emergencies (VII)	178	.80	.03	29.52	<u>Emergency maneuvers (XIII) and:</u>				
55. Solo (VIII)	93	.69	.06	12.55	83. Cross-Country (XIV-XV)	162	.79	.03	27.31
56. Precision landings (X-XII)	178	.78	.03	27.07	84. Test practice (XVI)	169	.75	.03	22.03
57. Emergency maneuvers (XIII)	178	.71	.04	18.55	85. Flight exam. by Instr. (XVII)	156	.70	.04	17.10
58. Cross-Country (XIV-XV)	162	.60	.05	12.04	<u>Cross-Country (XIV-XV) and:</u>				
59. Test practice (XVI)	169	.53	.06	9.73	86. Test practice (XVI)	154	.78	.03	24.41
60. Flight exam. by Instr. (XVII)	156	.51	.06	8.45	87. Flight exam. by Instr. (XVII)	142	.74	.04	19.42
					<u>Test practice (XVI) and:</u>				
<u>Stalls, Spins (VI) and:</u>									
61. Emergencies (VII)	171	.62	.05	13.13	88. Flight exam. by Instr. (XVII)	149	.79	.03	26.47
62. Solo (VIII)	92	.56	.07	7.75					
63. Precision landings (X-XII)	171	.46	.06	7.72					
64. Emergency maneuvers (XIII)	171	.44	.06	7.03					

TABLE VI-A

MANEUVER INTERCORRELATIONS IN THE ORDER OF THE "r" VALUE

(All ratings are instructors' mean ratings for students with no previous training)

	<u>r</u>		<u>r</u>
Stalls, Spins and Test practice	.29	Cross-country and Emergencies	.69
Stalls, Spins and Flight exam.	.33	Solo and Landings	.69
Stalls, Spins and Cross-country	.37	Solo and Take-off	.69
Stalls, Spins and Emergency man- euvers (Stalls, Spins, Slips)	.44	Flight exam. and Emergency maneuvers	.70
Solo and Cross-country	.45	Emergency maneuvers and Landings	.71
Solo and Flight exam.	.45	Emergency maneuvers and Take-off	.71
Stalls, Spins and Precision landings	.46	Flight exam. and Cross-country	.74
Solo and Test practice	.49	Test practice and Precision landings	.74
Flight exam. and Airwork	.50	Emergency maneuvers and Emergencies	.75
Test practice and Airwork	.50	Test practice and Emergency maneuvers	.75
Flight exam. and Landings	.51	Precision landings and Airwork	.76
Stalls, Spins and Take-off	.51	Precision landings and Cross- country	.77
Test practice and Landings	.54	Precision landings and Take-off	.78
Flight exam. and Take-off	.54	Test practice and Cross-country	.78
Solo and Emergency maneuvers	.54	Precision landings and Landings	.79
Stalls, Spins and Landings	.55	Emergency maneuvers and Cross- country	.79
Test practice and Take-off	.55	Flight exam. and Test practice	.79
Stalls, Spins and Solo	.56	Emergencies and Take-off	.80
Stalls, Spins and Airwork	.57	Emergencies and Landings	.80
Cross-country and Airwork	.57	Emergencies and Precision landings	.81
Test practice and Emergencies	.60	Emergencies and Airwork	.82
Cross-country and Landings	.60	Emergency maneuvers and Precision landings	.86
Flight exam. and Emergencies	.60	Landings and Airwork	.91
Stalls, Spins and Emergencies	.62	Take-off and Airwork	.92
Cross-country and Take-off	.62	Take-off and Landings	.93
Solo and Emergencies	.64		
Solo and Precision landings	.65		
Solo and Airwork	.68		
Emergency maneuvers and Airwork	.68		
Flight exam. and Precision landings	.68		

EACH MANEUVER LISTED WITH OTHER MANEUVERS IN THE ORDER OF THE "P" VALUE
AND WITH THE MORE GENERAL RATINGS OF THE TRAVERSE AND FLIGHT MEAN. 18

(All ratings are by instructors. Students considered are
those with no previous training.)

	<u>P</u>		<u>P</u>
<u>Take-off and:</u>		<u>Stalls, Spins and: (Cont'd)</u>	
Stalls, Spins	.51	Take-off	.51
Flight exam. mean	.54	Landings	.55
Test practice	.55	Solo	.56
Cross-country	.62	Airwork	.57
Solo	.69	Emergencies	.62
Emergency maneuvers	.71		
Precision landings	.78	<u>Emergencies and:</u>	
Emergencies	.80	Test practice	.60
Airwork	.92	Flight exam. mean	.60
Landings	.93	Stalls, Spins	.62
		Solo	.64
<u>Airwork and:</u>		Cross-country	.69
Flight exam. mean	.50	Emergency maneuvers	.75
Test practice	.50	Take-off	.80
Stalls, Spins	.57	Landings	.80
Cross-country	.57	Precision landings	.81
Solo	.68	Airwork	.82
Emergency maneuvers	.68		
Precision landings	.76	<u>Solo and:</u>	
Emergencies	.82	Cross-country	.45
Landings	.91	Flight exam. mean	.45
Take-off	.92	Test practice	.49
		Emergency maneuvers	.54
<u>Landings and:</u>		Stalls, Spins	.56
Flight exam. mean	.51	Emergencies	.64
Test practice	.54	Precision landings	.65
Stalls, Spins	.55	Airwork	.68
Cross-country	.60	Landings	.69
Solo	.69	Take-off	.69
Emergency maneuvers	.71		
Precision landings	.79	<u>Precision landings and:</u>	
Emergencies	.80	Stalls, Spins	.46
Airwork	.91	Solo	.65
Take-off	.93	Flight exam. mean	.68
		Test practice	.74
<u>Stalls, Spins and:</u>		Airwork	.76
Test practice	.29	Cross-country	.77
Flight exam. mean	.33	Take-off	.78
Cross-country	.37	Landings	.79
Emergency maneuvers	.44	Emergencies	.81
Precision landings	.46	Emergency maneuvers	.86

TABLE VI-B (CONTINUED)

	<u>r</u>		<u>r</u>
<u>Emergency maneuvers and:</u>		<u>Test practice and:</u>	
Stalls, Spins	.44	Stalls, Spins	.29
Solo	.54	Solo	.49
Airwork	.68	Airwork	.50
Flight exam, mean	.70	Landings	.54
Landings	.71	Take-off	.55
Take-off	.71	Emergencies	.60
Emergencies	.75	Precision landings	.74
Test practice	.75	Emergency maneuvers	.75
Cross-country	.79	Cross-country	.78
Precision landings	.86	Flight exam, mean	.79
<u>Cross-country and:</u>		<u>Flight exam, mean and:</u>	
Stalls, Spins	.37	Stalls, Spins	.33
Solo	.45	Solo	.45
Airwork	.57	Airwork	.50
Landings	.60	Landings	.51
Take-off	.62	Take-off	.54
Emergencies	.69	Emergencies	.60
Flight exam, mean	.74	Precision landings	.68
Precision landings	.77	Emergency maneuvers	.70
Test practice	.78	Cross-country	.74
Emergency maneuvers	.79	Test practice	.79

18 Editor's Note. It has been suggested that the correlations might possibly be attributed to two factors -- a general factor of flying proficiency, and a factor associated with the stage of training at which the observation is made. There is support for the suggestion in a comparison of the first six lists with the last five. The magnitude of the coefficients in the first six lists of intercorrelations between mean ratings is directly related to the reliability of the intercorrelated variables (if, as appears logical, the mean ratings of variables rated a great number of times are more reliable than are the mean ratings of variables rated only a few times). That is, intercorrelations between the mean ratings of two variables which were rated a great number of times, are greater than intercorrelations between mean ratings of one variable rated a great number of times, and another variable rated only a few times. Intercorrelations between two variables each rated only a few times are lowest.

In contrast, the intercorrelations in the last five lists are not patterned as to magnitude according to the reliability of the intercorrelated variables. The highest intercorrelations in each of these lists are between maneuvers or types of maneuvers which have not received large numbers of ratings, but are introduced as practice maneuvers in the final 8 or 10 hours of the course.

TABLE VII

INTERCORRELATIONS BETWEEN MEANS OF RATINGS GIVEN
DURING VARIOUS STAGES OF FLIGHT INSTRUCTION TO STUDENTS
HAVING PREVIOUS TRAINING

<u>No.</u>	<u>N</u>	<u>r</u>	<u>σr</u>	<u>$\frac{r}{\sigma r}$</u>
89. Stage A and Stage B	110	.64	.06	11.19
90. Stage A and Stage C	110	.53	.07	7.85
91. Stage B and Stage C	110	.79	.04	21.24
92. Stage A-1 and Stage A-2	110	.71	.05	15.19
93. Stage A-1 and Stage A-3	110	.65	.06	11.80
94. Stage A-1 and Stage A-4	110	.60	.06	9.63
95. Stage A-2 and Stage A-3	110	.77	.04	19.72
96. Stage A-2 and Stage A-4	110	.74	.04	16.70
97. Stage A-3 and Stage A-4	110	.86	.03	34.44

TABLE VII-A

COMPARATIVE LISTING OF CORRELATIONS
BETWEEN STAGES IN TRAINING FOR STUDENTS WITH AND
WITHOUT PREVIOUS TRAINING

	No previous training	With previous training
<u>Intercorrelations between stages in training</u>	<u>r</u>	<u>r</u>
Stage A and Stage B	.67	.64
Stage A and Stage C	.54	.53
Stage B and Stage C	.81	.79
Stage A-1 and Stage A-2	.71	.71
Stage A-1 and Stage A-3	.60	.65
Stage A-1 and Stage A-4	.44	.60
Stage A-2 and Stage A-3	.82	.77
Stage A-2 and Stage A-4	.64	.74
Stage A-3 and Stage A-4	.78	.86

Correlations for students with previous training. There were 110 students who had had previous training of some sort and they were eliminated from the study described above. This group of 110 had had dual time ranging from one-half hour to 26 hours, solo time ranging from a few minutes to 150 hours, and/or glider flights ranging from 2 to 1500. It was thought that some correlations on this group with the common factor of previous training, a factor absent from the group of 178 students considered in the first set of correlations, would show whether or not it would be advisable in future training to make separate groupings of students on the basis of this factor. The results are shown in Tables VII and VIII. In Tables VII-A and VIII-A comparisons are made between the "r" values for the two groups. The correlation coefficients have a fairly close correspondence to those for the group without previous training. In the correlations between inspectors' general ratings and the instructors' mean ratings on maneuvers, the "r" values are not only low but are also entirely insignificant.

Inspection shows that the "r" value in the correlation on Stage A-1 with Stage A-4 for the students who had had no previous training was .44, while for those with previous training it was .60. The latter appears to show closer agreement. However, we can set up an hypothesis that the two samples came from the same parent population and test the significance of the difference between the two correlations by use of the z-transformation. The difference in the z-values is .22; the standard error of the difference is .12; and diff. / σ diff. is only 1.80. Hence the difference is not significant.¹⁹ Therefore, the results are not inconsistent with the hypothesis. The two groups would not have had to be treated separately but could have been put together to make a larger frequency. Two other differences were tested, that between the correlations of A-2 and A-4, and between the correlations of A-3 and A-4. These results also showed insignificant differences.

Time in stages of training. On the forms used for the tabulation of averages space was provided for the time spent by the student on each stage and his total time; also, the time spent by each individual instructor on each stage and his total time; as well as his individual mean ratings. It was found that the number of students trained throughout the course by one instructor was too small to afford valid data for statistical treatment, so the plan to make some comparisons of individual instructors had to be abandoned. One instructor carried eight students throughout the whole course; four others trained seven students each. In all other cases the student was trained by different instructors numbering from two to five.

¹⁹ Editor's Note. It should be noted, however, that the fact that two correlation coefficients are not significantly different does not represent crucial evidence that the two samples yielding the coefficients are not from different populations, since nothing specific is indicated regarding the means and standard deviations of the two samples. The two correlations might, in fact, involve different aspects of a complex criterion.

TABLE VIII

**CORRELATIONS BETWEEN INSPECTORS' RATINGS AND INSTRUCTORS'
MEAN RATINGS GIVEN STUDENTS HAVING PREVIOUS FLIGHT TRAINING**

<u>Inspectors' rating and:</u>	<u>N</u>	<u>r</u>	<u>σr</u>	<u>$\frac{r}{\sigma r}$</u>
98. Groundwork (I) mean rating	110	.13	.09	1.40
99. Taxiing (II) mean rating	110	.20	.09	2.17
100. Take-off (III) mean rating	110	.20	.09	2.15
101. Airwork (IV) mean rating	110	.16	.09	1.68
102. Landings (V) mean rating	110	.15	.09	1.66
103. Stalls, Spins (VI) mean rating	107	.01	.10	0.08
104. Emergencies (VII) mean rating	110	-.01	.10	0.14
105. Solo (VIII) mean rating	55	.00	.14	0.01
106. Precision landings (X-XII) mean rating	110	.12	.09	1.26
107. Emergency maneuvers (XIII) mean rating	109	.08	.10	0.83
108. Cross-country (XIV-XV) mean rating	106	-.02	.10	0.18
109. Test practice (XVI) mean rating	106	.11	.10	1.10
110. Instr. Flight exam. (XVII) mean rating	93	.04	.10	0.41
111. C-VII to XVII mean rating	110	.05	.10	0.50
112. Instructors' general mean rating	110	.15	.10	1.59

TABLE VIII-A

**COMPARATIVE LISTING OF CORRELATIONS BETWEEN INSPECTORS'
RATINGS AND INSTRUCTORS' MEAN RATINGS FOR STUDENTS
WITH AND WITHOUT PREVIOUS TRAINING**

	No previous training	With previous training
<u>Inspectors' rating and:</u>	<u>r</u>	<u>r</u>
Groundwork	.17	.13
Taxiing	.20	.20
Take-off	.17	.20
Airwork	.22	.16
Landings	.21	.15
Stalls, Spins	.00	.01
Emergencies	.17	-.01
Solo	-.05	.00
Precision landings	.14	.12
Emergency maneuvers	.13	.08
Cross-country	.09	-.02
Test practice	.07	.11
Flight exam. by Instr.	.13	.04
C-VII to XVII	.17	.05
Instructors' general mean rating	.19	.15

Since it was specified on this program that the minimum dual time should be eight hours, the minimum time for Stage B six hours, the minimum time for Stage C twenty-one hours, and the minimum total time thirty-five hours, the treatment of time has no value as a measure of progress. In lieu of this, some time-data were used which had been accumulated at Kelly Field in 1918 by Dr. Henmon, who was doing some research work in aviation with Dr. Thorndike. These data gave the dual, primary, advanced, and cross-country and R. M. A. training time for 300 cadets. The results are shown in Table IX.

TABLE IX
CORRELATIONS IN TIME BETWEEN STAGES OF TRAINING
FOR 300 CADETS AT KELLY FIELD, 1918

	<u>r</u>	<u>σ_r</u>	<u>$\frac{r}{\sigma_r}$</u>
1. Dual with Primary	.10	.06	1.82
2. Dual with Advanced	.26	.05	4.80
3. Dual with C.C. and R.M.A. ²⁰	.23	.06	4.13
4. Primary with Advanced	.19	.06	3.43
5. Primary with C.C. and R.M.A.	.18	.06	3.16
6. Advanced with C.C. and R.M.A.	.43	.05	9.09

Means²¹

1. Dual	$\bar{X} = 9.66$ hrs., or 9 hrs. 40"
2. Primary	$\bar{X} = 11.56$ hrs., or 11 hrs. 34"
3. Advanced	$\bar{X} = 14.93$ hrs., or 14 hrs. 56"
4. C.C. and R.M.A.	$\bar{X} = 10.60$ hrs., or 10 hrs. 36"

"REMARKS" BY INSTRUCTORS

On each sheet of the records used for rating students there is a space provided, opposite each flight, for a remark by the instructor. To supplement the analysis of ratings, it was planned to make some bi-serial correlations, using the data in these "Remarks." Two classes of "tense" and "not tense" could be considered, or good and poor judgment, or even favorable and unfavorable. However, this plan was given up because it could not be assumed that the student was not tense in flights where no mention of "tense" or "at ease" was made.

In addition there were conveyed too many shades of meaning. For example, does nervousness include tenseness? Does "bad landing" necessarily mean poor judgment? It might include poor judgment as a factor but it might mean too scant a knowledge of or a poor memory of instruction.

Examination of the remarks showed also that favorable and unfavorable classes could not be set up because too many questionable elements entered into an interpretation of the remarks on this basis.

²⁰Commission as Reserve Military Aviator.

²¹In the statistical treatment, time was rounded off to the nearest half-hour.

It was then intended to make a tabular count of certain attributes and try to classify types of behavior, but again there were too many difficulties and too much uncertainty in interpreting the meaning of the remark in the mind of the instructor when he wrote it down. As this analysis would have no statistical significance, it was not carried out because the time necessary for its completion could not be justified. All that can be done then is to mention some obvious features to which attention is called by mere inspection and make some general criticisms.

In the first place, entirely too many remarks are noncommittal. "Solo 30° eights," "Solo practice on power turns," "Solo landings," "Practice on spins, stalls, slips," are a few examples of a large number of remarks that are merely unnecessary statements by the instructor that that type of practice was given attention in that particular period. They are unnecessary because the ratings given that maneuver in that period would suffice in giving the same information. Still more meaningless are "Solo practice," "Dual practice," and "Check flight." Also, many times the remarks are too meager and in no way analytical. "O.K." is a favorite expression and may occur as often as ten times on one page, in conjunction with approaches, turns, or whatever maneuver is judged "O.K."

On the other hand, many remarks are technical criticisms and give valuable and enlightening suggestions for the use of the student or of another instructor. Some examples of these are: "Undershooting due to going out too wide," "Drops all landings about 10 feet," "Lets nose get too low in turns in gliding," "Commences bank too early on 70° eights," "Overshoots when coming in from spinning," "Hesitates on pulling nose up to stall," etc.

Many remarks reveal the progress of the student in performance technique and his flying aptitude. They are illustrated by "Getting feel of ship," "Jerky on controls," "Steady improvement," "Poor conception of drift," "Too tense -- especially with feet," "Slow in emergencies," "Poor judgment of gliding distance," "Acts well in emergency." These too should be of value in a check-up by a student or another instructor.

The remarks that occur possibly the most often indicate the attitude of the student, and in the following year's program a column headed "Attitude" was provided for a rating. Some of these remarks are "At ease, calm, relaxed, cool-headed, self-confident, thoughtful, intelligent, apt, capable, figures things out, clear thinking, learns quickly, conscientious, careful, conservative, cautious, tries hard, natural, corrects his faults, precise, steady, willing and attentive to instructions, not apprehensive, acute and receptive mind, no bad habits, good reactions, (one remark stated 'no reactions')." Many of these remarks are also adverse criticism, such as "Cocky, excitable, over-anxious, indifferent, careless about regulations, can't be bothered, inconsistent, awkward, tense, nervous, too hurried, forgets small things, slow to respond, conceited, tendency to show off, self-

conscious with instruction, inattentive, incooperative but can't apply, tries too hard, bewildered, unobservant, lacks initiative, timid, not mechanically inclined."

A matter deserving the most severe criticism is the inconsistency of many of the remarks with the ratings given during the corresponding flight period. This inconsistency appears over and over again and immediately places under suspicion the reliability of all the remarks for any purpose whatsoever. Very little attention was paid by some instructors to the fact that the rating scale defines 1-good, 2-above average, 3-average, 4-below average, 5-poor. A remark will read, "Airwork unusually good," and a check of the grades given in that period shows four 2's under Airwork, four 2's under Taxiing, four 3's under Groundwork, one 1 under Take-off, and one 1 under Landing.

Following are a few samples of this constantly recurring type of inconsistency with the corresponding grades:

"Above average. Coordination good" -- all grades 3.

"Approaches and landings poor" -- grade 3.

"Did poor on 360" -- grade 3.

"Normal" -- grade 1.

"Did poor work today" -- grades 2's, 3's, 4's.

"Does not get tail down; balloons on landing" -- grade 1.

"Excellent landings" -- grade 3.

"Landings very good" -- grade 3.

"Spins very good" -- grade 2.

("All 8's very good" -- grade 2.

("All 8's much better" -- still graded 2.

"Approaches and landings poor" -- grade 4.

("Solo landings all very good" -- grade 2 on Stage B.

("Landings fair" -- grade 2 on Stage C.

Another variety of inconsistency is the noting of improvement in certain maneuvers while the grade has remained the same.

("Still not satisfied with landings" -- grade 3.

("Landings improved" -- grade 3.

("Landings O.K. now" -- still grade 3.

Here the instructor gives evidence of a serious endeavor to perfect the landing maneuver, but his rating is something entirely removed from that field of endeavor and is totally unrelated to the progress made by the student through the instructor's efforts. This recurs over and over again through the reports. A remark may say, "Slight relapse in distance judgment" -- but it had no effect on his grades. The remark may follow, "Distance judgment improved" -- but not his grades. One example, selected at random, reads, "General improvement. Above average student," yet all the grades are 3's identical with the grades given in the ten preceding flight periods. In fact, these grades remained the same throughout the fifteen periods of Stage B. The only encouragement to be derived by this student from the remark by the instructor concerning general improvement was the fact that at least his grades did not regress. The frequent recurrence of examples where improvement is noted without any improvement of grades leads inevitably to the conclusion that the remarks could not possibly have been filled in at the same time as the grades, but must have been set down subsequently or perhaps even previously.

Still another even more glaring inconsistency occurs all too often. This is a special remark about the practice of a maneuver or improvement in the performance of a maneuver, with no grade at all given under that maneuver, and is illustrated by the following:

"180° landings good" -- no grade.

"Loses altitude on 70° turns" -- no grade.

"Rights, spots, better" -- no grade.

"Precision landings improved" -- no grade.

"Holding altitude better" -- no grades for any maneuver.

"Good on 90° and 180°" -- no grade.

This of course again indicates clearly by such a lack of coincidence that there is room for improvement in the essential requirements of the present rating system.

Some curious things are noticed about the use of a rating scale of 1, 2, 3, 4, 5. A student will have received straight 1's in Groundwork and Taxiing through most of Stage A and all of Stage B. Then Stage C, when some advanced maneuvers are taken up and are graded 3, strangely enough the grades on Groundwork and Taxiing also become 3's. And why should a student who has been receiving straight 1's in take-offs, airwork, and landings suddenly be given straight 2's in those maneuvers? Moreover, they remained 2's for thirteen flight periods. A change of instructor would be suspected, but the same initials were signed throughout.

It is also noticed that there will be progress in grades from 3's to 1's on Groundwork, Taxiing, Take-off, Airwork, and Landings, but in precision work the grades do not progress.

Another curious example is that of a pendulum swing -- all 3's for two periods, all 2's for the next two, all 3's again for the next two periods, and all 2's again for the next two. It is humanly possible that this pendulum rating is indicative of a variation in judgment on the part of the instructor rather than a variation in performance on the part of the student.

The conclusion that even a casual reading of the "Remarks" compels is that if large numbers of the remarks are so conspicuously inconsistent with the grades for corresponding periods, or if large numbers of them are indeterminate in meaning, they serve no conceivable purpose. Both the system of rating and the requirement for explanatory remarks should at least be of such kind that their execution will not be too intricate or perplexing for the instructor's ability and not too involved for a reasonable expenditure of time devoted to them.