

THE DEVELOPMENT OF
"A SCALE FOR RATING PILOT COMPETENCY"

by

E. LOWELL KELLY

with an
Appendix

A Factor Analysis of the Purdue
"Scale for Rating Pilot Competency"

by

Robert J. Wherry and Robert C. Rogers

A report on research conducted at Purdue University, Lafayette, Indiana, 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.

July 1943

CIVIL AERONAUTICS ADMINISTRATION
Division of Research
Report No. 18
Washington, D. C.

PROCESSED BY
CIVIL AERONAUTICS
ADMINISTRATION
WASHINGTON, D. C.

53905
1943
JUL 11 1943
NATIONAL RESEARCH COUNCIL

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

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

M. S. Viteles, Chairman

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

July 28, 1943

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

Dear Dr. Brimhall:

Attached is a report entitled The Development of "A Scale for Rating Pilot Competency" by E. Lowell Kelly, with an Appendix by R. J. Wherry and R. C. Rogers entitled A Factor Analysis of the Purdue "Scale for Rating Pilot Competency." The report is submitted by the Committee with the recommendation that it be included in the series of technical reports issued by the Division of Research, Civil Aeronautics Administration.

In reading this report it should be borne in mind that it was originally written in 1940 and therefore expresses the point of view of the author as of that date. Nevertheless, the report is of current interest because it embodies considerable data concerning an instrument, "A Scale for Rating Pilot Competency", which has been extensively used in investigations on the selection and training of aircraft pilots sponsored by the Committee on Selection and Training of Aircraft Pilots.

Cordially yours,



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

MSV:es

Rec 8-16-43

EDITORIAL FOREWORD

In the research program sponsored by the Committee on Selection and Training of Aircraft Pilots, considerable use has been made of "A Scale for Rating Pilot Competency," developed by E. L. Kelly at Purdue University. The present report describes early studies which led to the development of the scale in its final form.

The study is published because it contains a description of the method employed in developing the scale and considerable data on inter-relationships among items not present in other reports so far included in the Technical Series published by the Division of Research, Civil Aeronautics Administration. It is believed that such information will be of considerable value to research workers in this area of investigation.

In reading the report, it should be borne in mind that it was originally written in 1940 and expresses the author's point of view as of that date.

CONTENTS

	Page
Editorial Foreword	v
Summary	ix
Need for A Scale for Rating Pilot Competency	1
Preliminary Rating Scale	1
Description of the Experimental 14-Item Rating Scale	2
Validation of the 14-Item Scale	4
Factor Analysis of the 14-Item Scale	4
Discussion	15
 Appendix: A Factor Analysis of the Purdue "Scale for Rating Pilot Competency", by Robert J. Wherry and Robert C. Rogers	 17
I. Introduction	19
II. Procedure and Results	19
III. Recommendation	23

SUMMARY

Because of the absence of suitable criteria of success in flying, there is a real need for an improved scale for rating pilot competency. A preliminary "man to man" graphic scale covering three traits (skill, emotional stability, and judgment) was devised and tried out but discarded because of excessively high intercorrelations between traits.

Several other scales were constructed and then the best points of each combined into a 14-item graphic scale for experimental study. The scale proved to be easy to use and has been shown to differentiate well between good and poor pilots. The intercorrelations of the traits were high enough to suggest the existence of common factors but not so high as to indicate an undue halo effect.

A factor analysis of the table of intercorrelations shows that three factors are sufficient to account for all intercorrelations of items in the scale. These seem to be: skill, judgment, and emotional control, listed in order of the extent to which the scale measures each. A scheme is suggested for combining the ratings so as to obtain approximate measures of these factors. These three factor scores intercorrelated with each other to the extent of .40, .41, and .58. They should constitute useful criteria both for future research and for practical evaluation of pilot competency.

DEVELOPMENT OF "A SCALE FOR RATING PILOT COMPETENCY"¹

NEED FOR A SCALE FOR RATING PILOT COMPETENCY

All workers on the problem of pilot selection have realized the need for a reliable and valid measure of success in flying. To be able to predict behavior one must first have some means of evaluating that behavior. Since the beginning of the Civilian Pilot Training program, the performance of flight students has been rated on a "1 to 5" scale for each period of instruction, but these ratings have been found to be of questionable value as measures of pilot competency. Although highly reliable, their validity seems to have been vitiated by one or both of the following factors:

(a) Variations in interpretations of the meanings of the scale values by various instructors. For example, some instructors interpret "1" to mean "the highest quality performance, of the sort rarely exhibited by student pilots," whereas other instructors interpret "1" to mean "excellent performance for the amount of previous training which the student has had."

(b) The ratings often are made at the end of a day during which the instructor has had eight or ten different students. Consequently he may confuse the performance of one student with that of another when the ratings are made. In filling out the required ratings at the end of the day or "during the next rainy period," the instructor is probably tempted to give a rating indicative of his general impression of the student. This general impression, of course, does not vary greatly from day to day.

In any event, in spite of the amazingly high reliability of these day-to-day ratings, they have been found to correlate but very slightly with other measures of pilot competency and there is little evidence of correlation with predictive measures.²

PRELIMINARY RATING SCALE

The first attempt to design a scale to secure more than a single measure of pilot skill resulted in the form shown as Exhibit I. As can be seen, it was a combination of the "man to man" and graphic types of scales and called for ratings by instructors on three aspects of the students' behavior: "skill," "emotional stability," and "judgment."

1

The author was assisted in this study by A.P. Johnson, E.S. Ewart, G.A. Satter, Lyle Tussing, and James Wright.

2

Editor's Note: The author is here referring to an index, and not to single maneuver ratings. An index obtained by summing the weighted ratings throughout the training period and dividing by the total number of ratings was found to have a reliability of .98. (The reliability was determined by correlating an index for odd half-hours of flight training with one for even half-hours, and correcting by the Spearman-Brown formula.) See:

Kelly, E.L. Analysis of Test Data on 1938-39 C.A.A. Students. (Purdue Research Foundation, Project No. 1: Progress Report, September, 1940.) Washington, D.C.: National Research Council Committee on Selection and Training of Aircraft Pilots, 1940. (Copy in Committee files.)

Each of the five instructors at Purdue was asked to rate each of his ten students on this scale on two different occasions. Using these preliminary data, rank-order intercorrelations were computed for the three traits, as well as measures of reliability from one rating period to the next. Although the reliability of such ratings was fairly high, the intercorrelations among the three scales ran so high as to indicate little or no differentiation on the part of the raters with respect to the three traits. It was therefore concluded that:

(a) These factors were so highly intercorrelated that there was no need for differentiating them; or,

(b) Instructors were unable to differentiate among these three aspects of pilot competency; or,

(c) The scale and its instructions were not adequate to secure such differentiation.

Because of the obvious weaknesses of the scale just described, a number of different experimental rating scales of various types was devised. Some were graphic, some were numerical, some were check lists, but all had one common characteristic: they sampled a larger number of aspects of pilot competency which made it possible to make the individual items of the scales somewhat more specific. After considerable discussion of the relative merits of each of the proposed scales, the 14-item graphic scale shown as Exhibit II was decided upon.

DESCRIPTION OF THE EXPERIMENTAL 14-ITEM RATING SCALE

As can be seen (Exhibit II), the experimental form of the scale is of the graphic type and contains 14 items. Each item or trait is introduced by a question designed to be as unambiguous as possible to the pilot instructor doing the rating. An attempt was made to use descriptive phrases as "steps" or "landmarks" on the scale, specifically applicable as answers to the question introducing the item and sufficiently varied from item to item to reduce the halo effect in the ratings.

In addition to the five descriptive phrases placed along each scale, the rater is given the privilege of checking a space entitled "no opportunity to observe." It was felt that the provision of this space would lessen the tendency of raters to make unreliable ratings on certain of the items.

The 14 items or questions included in the scale were finally chosen from a much more extensive list of characteristics differentiating good and poor pilots, a list which had been compiled on the basis of a number of extended discussions of the problem with flight instructors and other pilots. The 14 items eventually selected seemed to be those most essential to pilot competency and yet neither ambiguous nor too overlapping in nature. Naturally, however, zero intercorrelations among the items were not expected.

Item 14, it will be noted, is an "over-all rating" and was included to give the flying instructor an opportunity for an evaluation of the general competency of the student being rated.

Directions: This scale is designed to determine how well, at this stage of flight training, you can predict the eventual pilot competency of your students. Each of your students are to be rated on three characteristics, namely, skill (ability to handle a ship properly in flight maneuvers), emotional stability (ability to remain cool in emergency situations) and judgment and caution (tendency to fly in such a way as to minimize the number of risks taken).

First, beginning with the scale on skill, rank each of your students by writing in their last name, at the right of the scale below. If you are unable to distinguish between two candidates, write their names on the same line. Then proceed in the same fashion to rank them on emotional stability and judgment. Remember to rank the students in each case according to how you feel they will stack up at the end of their C-119 training, not according to their present characteristics. Further, as you know, a student who ranks high on skill may rank low on emotional stability and judgment.

SKILL

EMOTIONAL STABILITY

JUDGMENT

Name

Name

Name

EXCELLENT or top-notch pilot.

Best of Average

AVERAGE or typical private pilot.

Poorest of Av.

POOR or doubtful that he will ever be able to obtain private certificate.

VALIDATION OF THE 14-ITEM SCALE

One means of validating the scale seemed to be that of securing ratings by many different instructors of the best and poorest students trained by those instructors during the 1939-40 C.P.T. program. To secure these ratings the local C.P.T. directors in 130 institutions were sent sets of materials to be handed on to each of their flight instructors. In this manner each of 400 instructors in 130 institutions was asked to select the best and the poorest of the ten students trained by him and then to rate each of these pairs on the experimental form of the scale. Ninety-one of the 400 instructors to whom forms were sent actually rated their best and poorest students and returned the forms.

These 91 pairs of rating scales were scored on a "20-point scale"³ and distributions made of the ratings on each trait for groups of "best" and "poorest" student pilots. Figures 1 to 14 show graphically the resulting distributions of the scores. Curve "A" is based on the 91 "best" students; curve "B" on the 91 "poorest" students; and curve "C" is the distribution of ratings for the 110 typical C.P.T. students used in the factor analysis of the ratings described in the following section. The mean rating for each group is also shown on the charts.

A study of the figures shows immediately that certain of the items tend to differentiate between good and poor students much better than others. For example, there is relatively little overlapping of A and B distributions for Items 1, 4, 5, and 14. Other items, however, show a large amount of overlapping and much less difference between means. (See Items 3, 6, 8, 10, 11, and 12.)

The means, standard deviations, and critical ratios of the differences between the means for Groups A and B are shown in Table I. The data in Table I indicate that ratings of skill, (Item 1), speed of learning, (Item 4), and having the feel of the ship, (Item 5) are the most closely related in instructors' selections of "best" and "poorest" students.⁴

FACTOR ANALYSIS OF THE 14-ITEM SCALE

In order to determine the relationships between the items on the scale it was necessary to secure ratings by flight instructors on a relatively unselected group of pilots. To this end each of the flight instructors at Purdue, Indiana, Michigan, Ohio State, and two other large universities was asked to rate all ten of his trainees. In this manner it was hoped to secure some 400 ratings. Unfortunately, it was possible to obtain these ratings for only the Purdue and Michigan C.P.T. groups and a portion of the Ohio State and Indiana groups. In all, ratings were obtained for 110 students.

3

Item 10, it will be noted, is a two-ended item in that both ends are "poor." For this item, the scale runs from 0 to 20 and back to 0 on the other end.

4

The clear distinction between the groups on Item 14 might be expected in view of the over-all nature of that rating. It should be noted that all of the ratings satisfy the statistical test for significance of difference between means of the group, in that for each rating the difference is more than three times the standard error of the difference.

A SCALE FOR RATING PILOT COMPETENCY

No opportunity to observe

1. Considering his training, how skillful is he in carrying out precision maneuvers (spot landings, figure eights, etc.)?

very skillful | high average | average | low average | very poor

2. How does he handle the controls?

greatly over or under controls | considerably over or under controls | some over or under control | handles controls fairly smoothly | very smoothly and correctly

3. How carefully does he check his plane and engine before taking off?

very carefully | carefully | reasonably carefully | not carefully enough | does not check it

4. As compared with the other students you have trained, how readily does he "catch on" to your instructions?

very fast learner | fast | average | slow | very slow

5. To what extent does he have the feel of a ship?

unusually well | well | fairly well | poorly | not at all... flies mechanically

6. Does he show respect for a ship and its motor?

takes excellent care | takes good care | shows reasonable respect for both | tends to be careless | no regard at all

7. How tense or relaxed is he when flying?

extremely tense | rather tense | slightly too tense | almost sufficiently relaxed | ideally relaxed

8. Is he inclined to show off while flying a plane?

almost always | frequently | sometimes | seldom | never

9. How easily does he become upset when something goes wrong, for example, a motor failure?

very easily upset | easily upset | sometimes upset | usually calm and controlled | always calm and controlled

10. How confident is he of his flying ability?

much too confident | slightly over confident | sensibly confident | not confident enough | entirely lacking in confidence

11. Does he like to try out new things, new maneuvers and cross country trips, for example?

always tries new things | frequently trying new things | sometimes tries new things | rarely tries new things | never tries new things

12. How good is his judgment with regard to taking flying risks? (weather, stunting, etc.)

extremely cautious takes no unnecessary risks | rarely uses poor judgment | takes some unnecessary risks | takes many unnecessary risks | extremely reckless

13. How well is he satisfied with his flying ability?

always tries to improve | considerable effort at improvement | some effort at improvement | fairly well satisfied | entirely satisfied

14. In your opinion, considering skill, emotional stability, judgment, etc., how good an "all-around pilot" is he likely to become?

top notch private pilot | better than average private pilot | average private pilot | poorer than average private pilot | very poor--will not fly long

Rated by _____ Instructor

A SCALE FOR RATING PILOT COMPETENCY

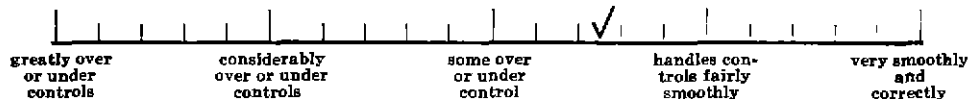
Directions for Use:

Being a good all-around pilot is not simply a matter of how well one can maneuver a ship, but also depends on the use of good judgment, on keeping one's head in emergencies, and on other traits which are difficult to measure. We have devised the scale on the back of this sheet especially to aid you in giving a many-sided description of a pilot and his flying habits.

The scale consists of 14 questions about the pilot being rated but a flat "Yes" or "No" answer is not asked for . . . rather you are given the opportunity of answering each question by checking at that point on the descriptive scale which best fits the pilot being rated. The descriptive phrases below each line should be thought of as "landmarks." Feel free to check any student as falling somewhere between these phrases if he belongs there.

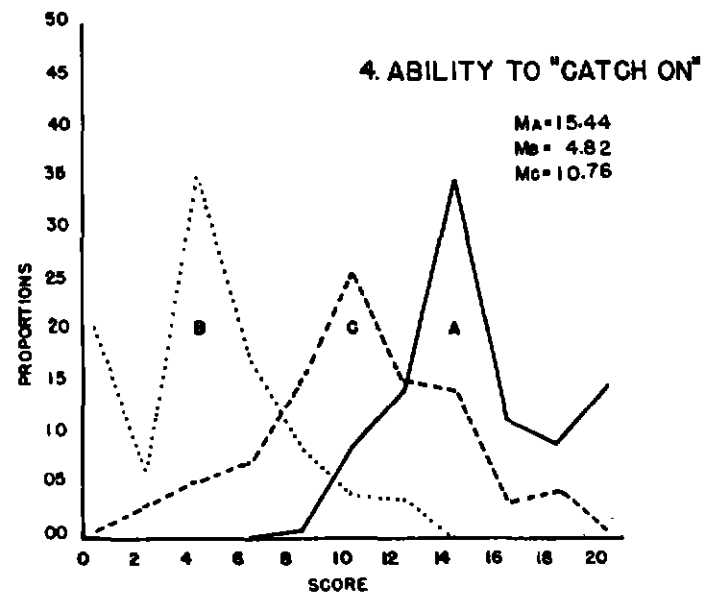
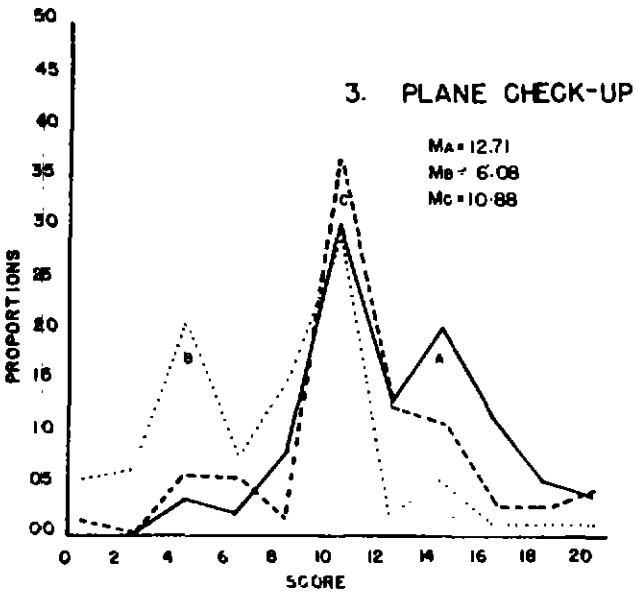
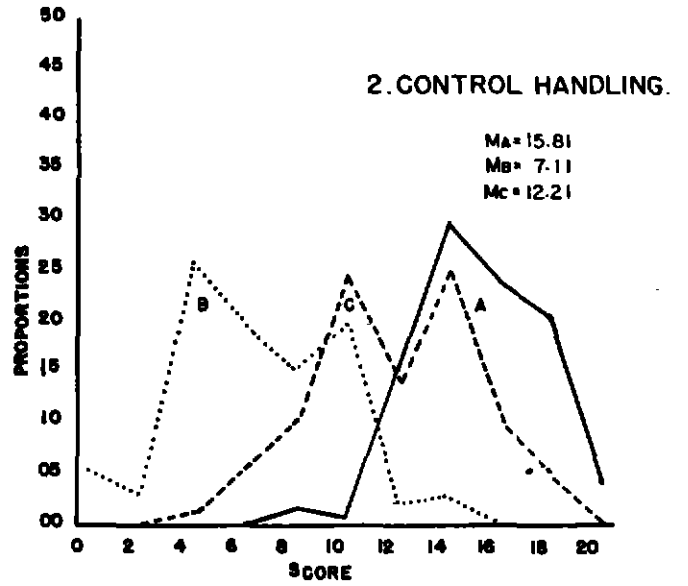
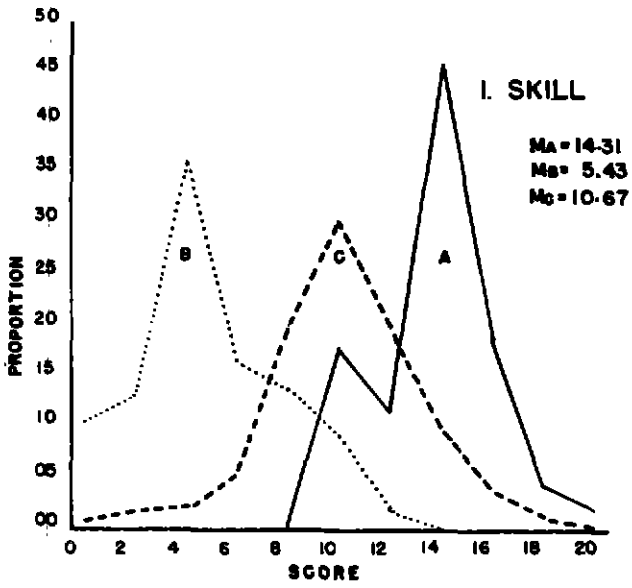
EXAMPLE: If, for example, a pilot usually handles the controls fairly smoothly but sometimes slightly over or under controls, you might rate him as indicated by the check mark on the sample scale below.

How does he handle the controls?



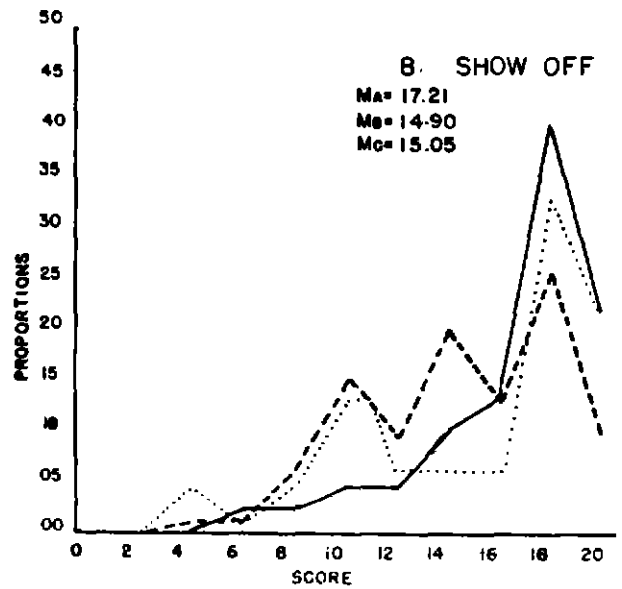
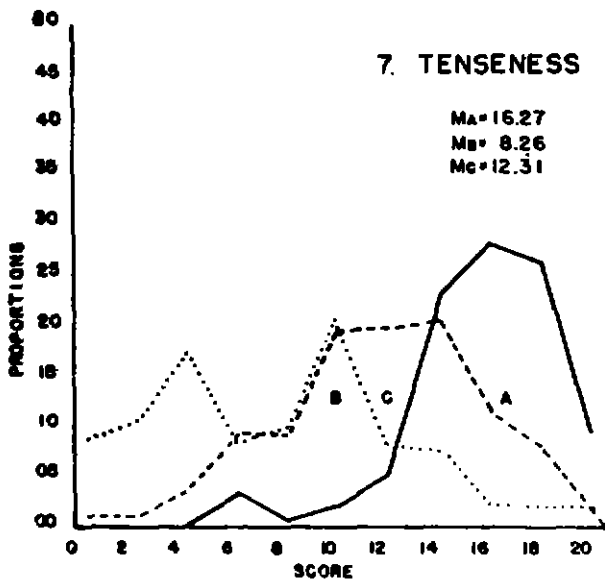
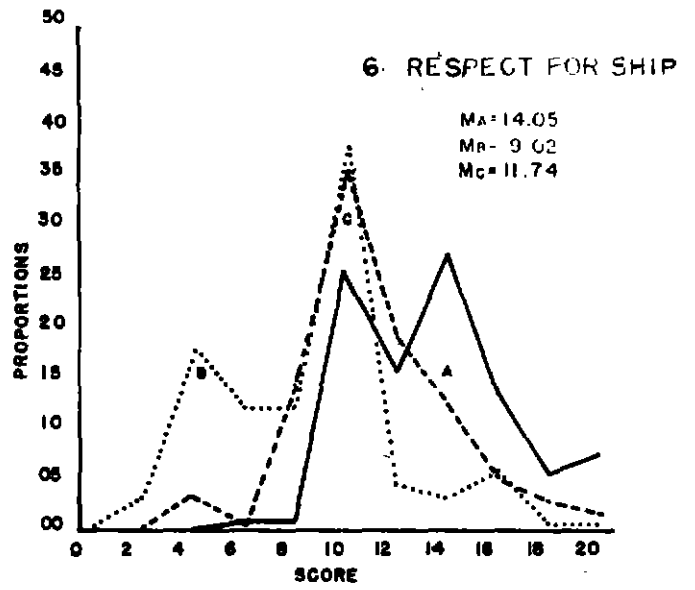
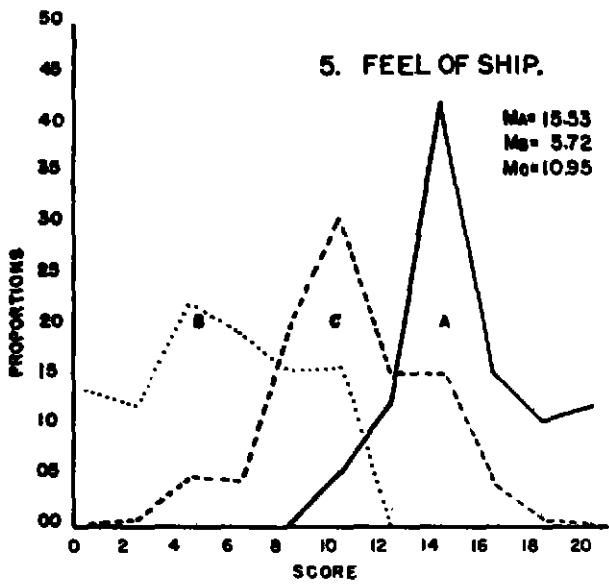
It should be remembered that just because a pilot deserves to be rated high or low on one characteristic does not mean that he deserves to be rated equally high or low on others.

PURDUE RESEARCH FOUNDATION



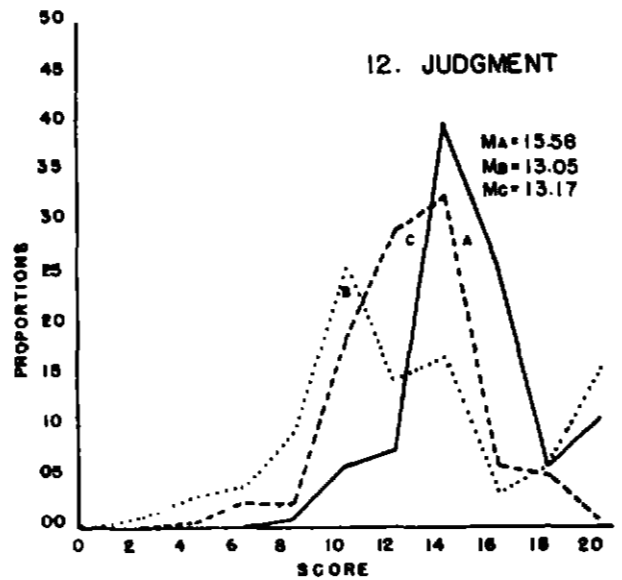
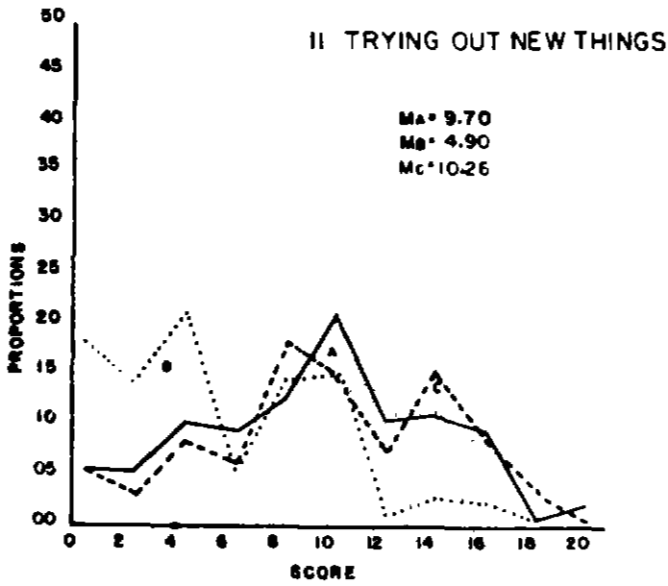
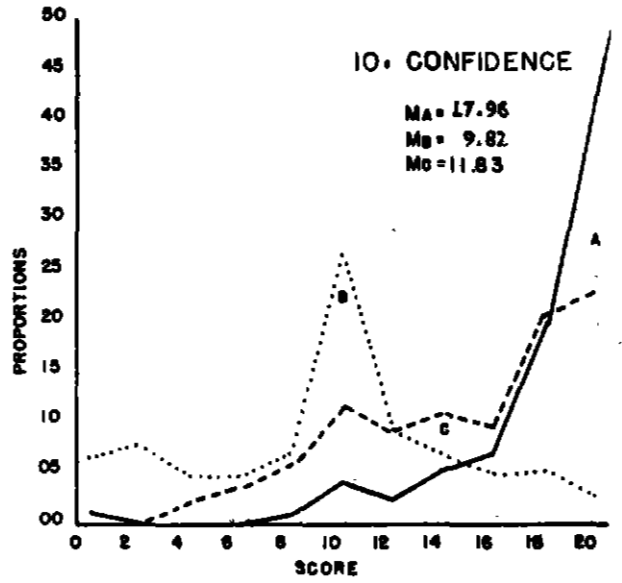
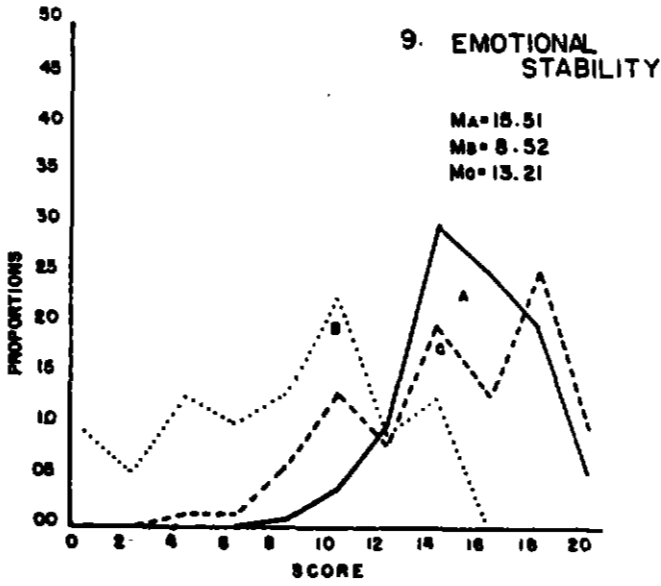
FIGURES 1 - 4

(See explanation in text)



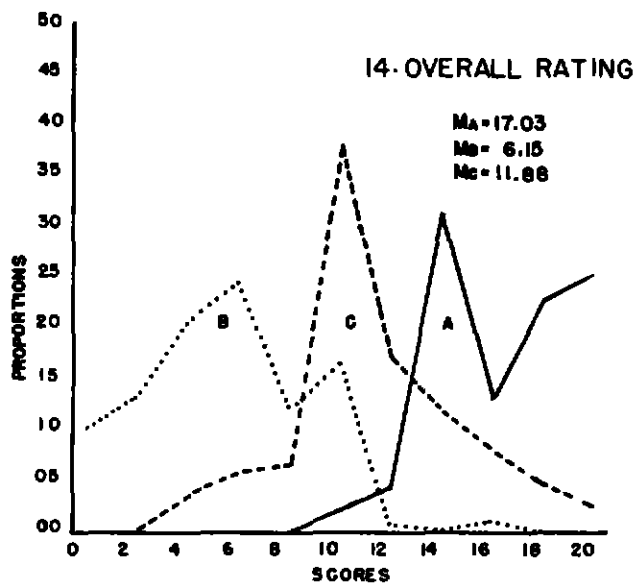
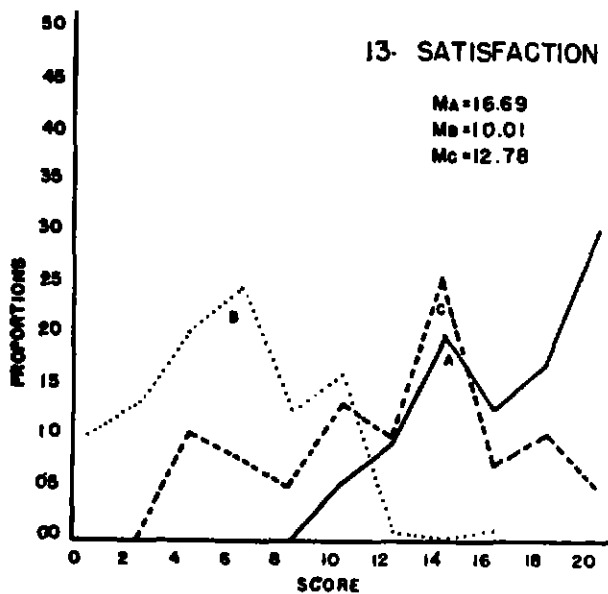
FIGURES 5 - 8

(See explanation in text)

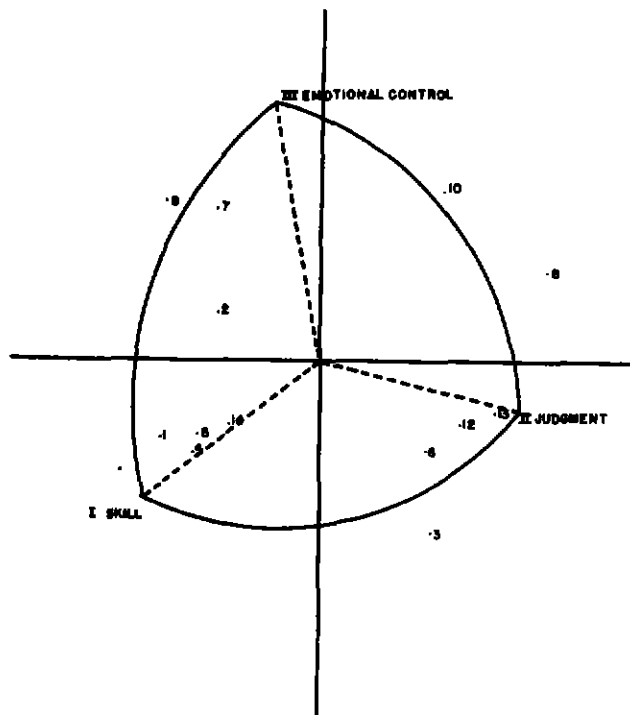


FIGURES 9 - 12

(see explanation in text)



15. FACTORIAL PATTERN OF THIRTEEN ITEMS ON THE PURDUE RATING SCALE



FIGURES 13 - 15

(See explanation in text)

TABLE I

Means and Standard Deviations of the Ratings of Groups A and B on
14 Items of Purdue "Scale for Rating Pilot Competency";
Also Differences of Means and Critical Ratios
for Each Difference

Item	Mean Group A	Mean Group B	diff. (A-B)	\bar{A}	\bar{B}	diff.	C.R.
1	14.31	5.43	8.88	2.40	2.84	.40	22.2
2	15.81	7.11	8.70	2.64	3.28	.44	19.8
3	12.71	6.08	6.63	3.57	5.15	.66	10.0
4	15.44	4.82	10.62	2.95	3.06	.48	22.1
5	15.53	5.72	9.81	2.53	3.14	.42	23.4
6	14.05	9.02	5.03	3.27	3.55	.50	10.1
7	16.27	8.26	8.01	2.81	4.97	.60	13.4
8	17.21	14.90	2.31	3.29	4.87	.62	3.7
9	15.51	8.52	6.99	3.08	4.50	.60	11.7
10	17.96	9.82	8.14	3.22	4.95	.61	8.1
11	9.70	6.50	3.20	5.53	4.90	.81	4.0
12	15.58	13.05	2.53	2.57	3.36	.46	5.5
13	16.69	10.10	6.59	3.30	5.29	.66	10.0
14	17.03	6.15	10.88	2.43	3.20	.42	25.9

These 110 scales, representing the ratings for a roughly random sampling of the C.P.T. students, were scored on the 20-point scale. As has been noted, the resulting distributions of these ratings are shown as graph "C" on Figures 1 to 14.

In practically all instances each instructor felt capable of rating a student on all 14 items. The one exception was Item 11, "Does he like to try out new things, new maneuvers and cross country trips, for example?" Only 92 of 110 scales included ratings on this item. The hesitancy which some instructors feel with regard to rating this item probably results from the fact that the C.P.T. program is fairly rigidly controlled and does not give a student much opportunity to reveal his tendency to experiment while participating in the course.

In order to determine the amount of interrelationship between the items and also to permit making a factor analysis of the ratings, intercorrelations were computed among all 14 items. These are shown in Table II. Rather than attempt a discussion of the correlation matrix at this point it seems desirable first to report the results of the factor analysis. Because of the smaller number of ratings available for Item 11 and the indication that this factor was not easily rated by C.P.T. instructors, Item 11 was omitted in the factor analysis. The remaining 13 variables were subjected to a factor analysis using Thurstone's Centroid Method.⁵

Three factors were found necessary to account for the intercorrelations

TABLE II

INTERCORRELATIONS OF SCORES OF GROUP C ON THE FOURTEEN ITEMS OF THE PURDUE[®]SCALE FOR RATING PILOT COMPETENCY[®] (N = 110)

Item No.	Item No.													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Skill	---	.647	.344	.792	.767	.365	.460	-.150	.358	.167	.380	.262	.127	.801
2. Control handling	---	---	.229	.576	.726	.464	.678	.156	.405	.225	.290	.377	.275	.595
3. Plane check-up	---	---	---	.330	.357	.590	.083	.084	.027	.004	.243	.470	.421	.427
4. Ability to "catch on"	---	---	---	---	.737	.420	.419	-.077	.391	.209	.356	.341	.253	.737
5. Feel of ship	---	---	---	---	---	.552	.492	-.066	.375	.117	.466	.327	.289	.752
6. Respect for ship	---	---	---	---	---	---	.307	.311	.168	.225	.230	.533	.463	.532
7. Tenseness	---	---	---	---	---	---	---	.167	.611	.270	.420	.172	.206	.422
8. Show-off	---	---	---	---	---	---	---	---	.123	.411	.331	.377	.318	.018
9. Emotional Stability	---	---	---	---	---	---	---	---	---	.119	.157	.234	.136	.364
10. Confidence	---	---	---	---	---	---	---	---	---	---	.151	.214	.241	.232
11. Trying out new things	---	---	---	---	---	---	---	---	---	---	---	.052	.323	.265
12. Judgment	---	---	---	---	---	---	---	---	---	---	---	---	.502	.473
13. Satisfaction	---	---	---	---	---	---	---	---	---	---	---	---	---	.363
14. Over-all	---	---	---	---	---	---	---	---	---	---	---	---	---	---

TABLE III

FACTOR LOADINGS (AFTER ROTATION) FOR THE THREE FACTORS ON THE 14 ITEMS OF THE PURDUE[®]SCALE FOR RATING PILOT COMPETENCY[®]

Factor Loadings

Item	I	II	III
1.	.90	-.06	.02
2.	.72	.16	.38
3.	.42	.50	-.36
4.	.85	.09	.01
5.	.87	.08	.03
6.	.47	.63	-.05
7.	.52	.06	.65
8.	-.15	.58	.37
9.	.45	.01	.48
10.	.10	.36	.36
12.	.34	.65	.01
13.	.24	.63	.02
14.	.86	.24	-.03

shown in Table II. Since the Thurstone technique does not yield psychologically meaningful factors until after the axes have been rotated, we shall not consider the original factor loadings but turn directly to the factor loadings after rotation. The loadings of each item on the three factors is shown in Table III. A factor loading, it will be remembered, may be interpreted as the coefficient of correlation between a given variable and the hypothetical factor. The Thurstone technique yields orthogonal (i.e., uncorrelated) factors, but in this problem it was found that the clusters of variables were not entirely uncorrelated, hence it was not possible to secure simple structure, i.e., entirely "pure" factors, by any rotation of the axes.

Factor I, it will be noted from Table III, shows a very high correlation with (i.e., is heavily saturated with) Items 1, 4, 5, and 14 on the scale, and a fairly high correlation with Item 2. Reference to the scale shows the items to be (here listed in order of decreasing size of the factor loadings):

1. "Considering his training, how skillful is he in carrying out precision maneuvers (spot-landings, figure eights, etc.)?"
5. "To what extent does he have the feel of a ship?"
14. "In your opinion, considering skill, emotional stability, judgment, etc., how good an 'all-around pilot' is he likely to become?"
4. "As compared with the other students you have trained, how readily does he 'catch on' to your instructions?"
2. "How does he handle the controls?"

Because of the nature of these items, it seems reasonable to name this first factor the "skill" factor.

Factor II, on the other hand, is less highly correlated with any of the items on the scale but from the relative size of the loadings will be seen to be most closely associated with Items 3, 6, 8, 12, and 13. These are (again in order of decreasing size of loadings):

12. "How good is his judgment with regard to taking flying risks (weather, stunting, etc.)?"
6. "Does he show respect for a ship and its motor?"
13. "How well is he satisfied with his flying ability?"
8. "Is he inclined to show off while flying a plane?"
3. "How carefully does he check his plane and engine before taking off?"

From the nature of these items, it seems that Factor II might well be named "judgment." It will be noted that the variables contributing to Factor I have very low loadings on Factor II and vice versa.

Factor III shows still lower correlations with any of the items. The items having the highest loadings are 2, 7, 9, and 10, which, in order of size of loadings

are:

7. "How tense or relaxed is he when flying?"
9. "How easily does he become upset when something goes wrong, for example, a motor failure?"
2. "How does he handle the controls?"
10. "How confident is he of his flying ability?"

Since these items all seem to involve an emotional component, this factor was tentatively named "emotional control."

The three factors and the clustering of the items which contribute most heavily to each can perhaps be better envisaged by reference to Figure 15 showing graphically these relationships when plotted on a spherical triangle. Here the clustering of Items 1, 4, 5, and 14 at the lower left corner is very obvious, and those in turn are some 75° away (90° separation equals zero intercorrelation) from the less compact cluster of 3, 6, 8, 12, and 13 at the lower right corner of the triangle. The lower weightings of 2, 7, 9, and 10 on Factor III are indicated by the distance which they are away from the apex of the triangle. That factors I and II are involved in Items 2, 7, 9, and 10 is indicated by the less than 90° separation of these points from the other corners of the triangle.

It appears, then, that the Purdue "Scale for Rating Pilot Competency" is in reality measuring three relatively disparate aspects of pilot competency: skill, judgment, and to a lesser degree emotional control. Rather than treating each of the 14 ratings as separate scores, it seems advisable to combine the scores by adding the ratings for each of the items in a cluster. Thus: for skill add Items 1, 4, 5, and 14; for judgment, Items 3, 6, 8, 12, and 13; and for emotional control, Items 2, 7, 9, and 10. Such combined scores will probably be much more reliable than ratings on single items. The intercorrelation of the three sums will not be zero, but will be sufficiently low so that the three combined scores may be treated as separate criteria both for research and for practical evaluation of pilot competency. Table IV shows the actual intercorrelations among the sums of ratings for the 110 C.P.T. students. As will be seen in Table IV, these sums which may be roughly considered as "factor scores" correlate with each other, .40, .41, and .58.

TABLE IV

	Judgment 3+6+8+12+13	Emotional Control 2+7+9+10
Skill 1+4+5+14	+ .40	+ .58
Judgment 3+6+8+12+13		+ .41

-25-

DISCUSSION

This scale has been shown to be statistically meaningful and practical to use inasmuch as instructors find it easy to rate students on it. No serious weaknesses have been found in the present form of the scale so no immediate revision seems necessary.

Eventually the reliability of the scale from rater to rater should be studied but since all C.P.T. students are trained by a single instructor, such a check has not been possible. It would also be desirable to have the same instructor rate students at different times to check the temporal reliability of the scale.

The data reveal nothing about the relative importance of these three factors in pilot competency. Factor I, skill, seems to be considered most important in the average instructor's judgment when he is asked to pick out his "best" and "poorest" pilots. Also, it will be noted that Item 14, a supposedly "all-around" judgment of pilot competency, is loaded much more heavily with the skill factor than with either judgment or emotional control.

It is conceivable that different types of flying call for different amounts of these three components of pilot competency. The "fair weather airport private pilot" probably needs more "skill" than anything else. If he attempts cross country alone, he needs "judgment" as well, and if his motor misbehaves, he needs a great deal of "emotional control." A pilot, on the other hand, might possibly have too much judgment to take certain types of risk which are necessary, for instance, in military flying. The "ideal" pilot would, of course, have optimum proportions of all these factors, but what constitutes the minimum amount of each factor and the optimum proportion of each must await future research.

In the meantime, however, it seems pertinent to raise the question of whether sufficient attention is being given to factors other than "skill" in the selection and training of student pilots. A certain degree of skill is obviously necessary before the other factors of judgment and emotional control can operate, but again it must be remembered that even skillful pilots may be unsafe pilots unless their skill is used with proper judgment and is unimpeded by a lack of emotional control. It has already been noted that instructors select their "best" students primarily on the basis of skill. Perhaps this is because of the limited amounts of training which their students had had at the time of the rating. It is possible, however, that this emphasis on skill is the result of our lack of knowledge of how to train students in such a way as to improve their judgment and emotional control -- or perhaps instructors consider these to be inborn traits of the student not susceptible to training. These questions suggest a number of fruitful researches in which the Purdue "Scale for Rating Pilot Competency" might be used as an evaluative device.

APPENDIX

A Factor Analysis of the Purdue[®]Scale for Rating
Pilot Competency"

by

Robert J. Wherry

and

Robert C. Rogers .

A progress report on research conducted at the University of North Carolina, Chapel Hill, North Carolina, 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.

A FACTOR ANALYSIS OF THE PURDUE "SCALE FOR RATING PILOT COMPETENCY"

I. INTRODUCTION

In his report on the construction of the Purdue "Scale for Rating Pilot Competency",¹ Kelly presented the results of a factor analysis of the scale which yielded three primary factors: Factor I:- Skill, was measured by Items 1,5,14,4, and 2; Factor II:- Judgment, by Items 12,6,13,8, and 3; and Factor III:- Emotional Control, by Items 7,9,2, and 10. However, in this preliminary study Kelly made no attempt to analyze the factors further in terms of isolating the best weighting procedures to be employed for these items in order to provide pure measures of the three factors assessed by the scale.

The purpose of the present study was to subject Kelly's data to further statistical treatment with the aim of providing weights for the various items which would permit the derivation of pure measures of the three factors. The data reported in this investigation provide also a cross-check on the stability of the factors derived by Kelly in his study.

II. PROCEDURE AND RESULTS.

Analysis of Kelly's original data. The first step in the statistical analysis was to subject Kelly's table of intercorrelations to an independent factor analysis and to re-rotate the original factor loadings.² This table of inter-

TABLE I

Intercorrelations of Scores of Kelly's Group C on the Fourteen Items of the Purdue "Scale for Rating Pilot Competency" (N = 110)

Item No.	Item No.													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Skill	---	647	344	792	767	365	460	-150	358	167	380	262	127	801
2. Control handling	---	---	229	576	726	464	678	156	405	225	290	377	275	595
3. Plane check-up	---	---	---	380	357	590	088	084	027	004	243	470	421	427
4. Ability to "catch on"	---	---	---	---	737	420	419	-077	391	209	356	341	253	737
5. Feel of ship	---	---	---	---	---	552	492	-066	375	117	466	327	289	752
6. Respect for ship	---	---	---	---	---	---	307	311	168	225	230	533	468	532
7. Tenseness	---	---	---	---	---	---	---	167	611	270	420	172	206	422
8. Show-off	---	---	---	---	---	---	---	---	123	411	331	377	318	018
9. Emotional Stability	---	---	---	---	---	---	---	---	---	119	157	234	136	364
10. Confidence	---	---	---	---	---	---	---	---	---	---	151	214	241	232
11. Trying out new things	---	---	---	---	---	---	---	---	---	---	---	052	323	265
12. Judgment	---	---	---	---	---	---	---	---	---	---	---	---	502	473
13. Satisfaction	---	---	---	---	---	---	---	---	---	---	---	---	---	368
14. Over-all	---	---	---	---	---	---	---	---	---	---	---	---	---	---

¹ A description of this scale and the original statistical analysis are presented in the body of the present report by E.L. Kelly.

² Thurstone, L.L. The Vectors of Mind. Chicago: University of Chicago Press, 1935.

correlations and the resulting factor loadings after rotation are presented in Tables I and II. It will be noted that the rotation carried out in the present study isolated practically the same factors as were reported by Kelly in his analysis. These factors have been given the same names as originally suggested by Kelly, i.e., Skill, Judgment, and Emotional Control.

TABLE II
FACTOR LOADINGS (AFTER ROTATION)
(N = 110)

Rating Items	WHERRY'S ROTATION			KELLY'S ROTATION		
	Skill	Emotional Control	Judgment	Skill	Emotional Control	Judgment
1. Skill in maneuvers	83	30	-21	90	02	-06
2. Handling of controls	60	57	09	72	38	16
3. Check plane	60	25	37	42	-36	50
4. Readiness in catching on	80	26	-05	85	01	09
5. Have feel of ship	82	30	-06	87	03	08
6. Respect of ship and motor	58	05	53	47	-05	63
7. Relaxed when flying	31	77	06	52	65	06
8. Non show-off	-12	25	65	-15	37	53
9. Calm in difficulty	29	60	00	45	48	01
10. Confident of ability	07	37	35	10	36	36
12. Judgment in taking risks	46	06	58	34	01	65
13. Tries to improve self	35	03	58	24	02	63
14. All-around pilot	87	21	09	86	-03	24

On the basis of this analysis, Items 1, 4, 5, 7, 8, 9, 12, 13, and 14 were selected for further use in scoring the scale. Item 2 was discarded because it was weighted too heavily on both skill and emotion; Items 3 and 6 because they were too highly weighted on skill and judgment; and Item 10 because its loadings were very low and split between emotion and judgment.

Analysis of North Carolina Data. The "Scale for Rating Pilot Competency" was administered to a new population of 136 cases. This population was made up of the following subsamples:

COLLEGE	NO. OF SUBJECTS
Washburn College	41
North Carolina State College	41
St. Frances College	14
Mass. State Teachers College	10
University of Wichita	24
University of North Carolina	6

Presented in Table III are the results of the factor analysis using only the items found to have relatively pure factor loadings in the above analysis of the Kelly data (Items 1, 4, 5, 7, 8, 9, 12, 13, and 14).

Table III

INTERCORRELATIONS (N = 136)

ROTATED (ORTHOGONAL)
FACTOR LOADINGS

Scale Item									ROTATED (ORTHOGONAL) FACTOR LOADINGS		
	1	4	5	7	8	9	13	14	Skill	Emotional control	Judgment
1.		.774	.773	.553	.152	.122	.466	.726	90	10	-.05
4.			.721	.583	.120	.524	.432	.722	85	16	.02
5.				.582	.219	.512	.492	.753	36	18	.07
7.					.152	.676	.323	.604	59	61	.02
8.						.070	.478	.306	20	03	.58
9.							.329	.555	51	58	.03
13.								.636	54	00	.62
14.									83	16	.27
12 ³	.26	.27	.23	.10	.60	.05	.52	.43	30	-.17	.75

Comparison of three scoring techniques. Presented in Tables IV, V, and VI are the intercorrelations between the factor scores when the items on the scale are scored by three different scoring (weighting) techniques.

Table IV presents the intercorrelations of the factors when 13 of the items are scored according to Kelly's original recommendations. In this scoring procedure, the ratings on items with high loadings on the separate factors are merely added together without weighting. In this case the Judgment score is obtained by summing items 3 + 6 + 8 + 12 + 13; the Emotional Control score = 2 + 7 + 9 + 10; and the Skill score is derived by summing items 1 + 4 + 5 + 14.

TABLE IV

INTERCORRELATIONS OF FACTORS BY KELLY'S SCORING METHOD (13 items;
(N = 136)

	<u>Judgment</u>	<u>Emotional Control</u>
<u>Skill</u>	.56	.72
<u>Judgment</u>		-.08

³ In this analysis Item 12 was originally omitted because of its high loading on "skill" in Table II, but when added here by Dwyer's extension it was found to be as satisfactory a measure of judgment as either Item 8 or 13 and therefore was retained. (Dwyer, P.S. The determination of the factor loadings of a given test from the known factor loadings of other tests. Psychometrika, 1937, 2, 173-178.)

⁴ Editor's Note: Kelly found that the combined ratings correlates with each other to the extent of .40, .41, and .53. (See Table IV, page 14.)

Table V presents the intercorrelations of the factors using 9 items. In this method only those items with relatively pure factor loadings were employed. These items are weighted by means of the actual regression coefficients found by calculating multiple correlations using the factor loadings as the criterion correlations. On this basis the three factors are scored in the following manner: Judgment = Item 8 + 4 x Item 12 + 2 x Item 13; Emotional Control = Item 7 + Item 9; and the Skill score = 3 x Item 1 + 2 x Item 4 + 2 x Item 5 + Item 14. The resulting factor intercorrelations are:

TABLE V
INTERCORRELATIONS OF WEIGHTED FACTOR SCORES (9 ITEMS USED)
(N=136)

	<u>Judgment</u>	<u>Emotional Control</u>
Skill	.44	.67
Judgment		.14

Intercorrelations between the predicted scores of orthogonal factors such as those reported in Table V are due to another type of difficulty, soluble only by the partialling techniques employed here. While the factors are orthogonal, the experimental variables used to define the factors are complex and contain overlap of the factors. It is this condition which accounts for the correlation between predicted factor scores of supposedly orthogonal factors. An inspection of the factor loadings in Table III and the intercorrelations in Table V indicates that:

1. The tests defining the skill factor are relatively pure measures of skill;
2. The tests defining the emotional control and judgment factors are pure as regards each other but all contain substantial amounts of the skill factor;
3. The purity of the skill factor makes it possible to partial the overlap of this factor from the other factor scores, thus leaving pure measures of emotional control and judgment as indicated in Tables VI and VII below.

Table VI presents the intercorrelations of the factors when the weighting procedures are further refined by partialling out the Skill from Judgment and Emotional Control. The factors then are scored by weighting the items in the following manner:

$$\begin{aligned} \text{Skill (SK)} &= 3 \times \text{Item 1} + 2 \times \text{Item 4} + 2 \times \text{Item 5} + \text{Item 14}. \\ \text{Judgment (J)} &= \text{Item 8} + 4 \times \text{Item 12} + 2 \times \text{Item 13} - .1783 \cdot \text{SK}. \\ \text{Emotional Control (E)} &= \text{Item 7} + \text{Item 9} - .3362 \cdot \text{SK}. \end{aligned}$$

TABLE VI
INTERCORRELATIONS OF WEIGHTED FACTOR SCORES (CORRECTED)

	<u>Judgment</u>	<u>Emotional Control</u>
Skill	.01	.01
Judgment		-.14

Table VII illustrates the fact that the factor loadings on corrected scores become relatively pure when items are weighted in the proper manner. For this table the factor scores were corrected on the basis of the new weights (partialling out skill as in Table VI) and the new loadings determined by Dwyer's Extension.⁵

TABLE VII
FACTOR LOADINGS IN CORRECTED SCORES (N = 136)
(Dwyer's Extension)

Factor	Skill Score	Judgment Score	Emotional Control Score
Skill	1.02	-.06	-.13
Judgment	.02	1.11	-.11
Emotional Cont.	.08	-.11	1.03

III. RECOMMENDATION

It is recommended that the present method of scoring be used in subsequent samples as a cross-check on the validity of factor scores as possible predictors of future flight success.

⁵ Scoring weights were derived by means of the following formulae:
1. $E \text{ corrected} = E - E_{SK}$ and $J \text{ corrected} = J - J_{SK}$, where:

$$E_{SK} = r_{E-SK} \frac{\sigma_E}{\sigma_{SK}} \cdot SK \text{ and } J_{SK} = r_{J-SK} \frac{\sigma_J}{\sigma_{SK}} \cdot SK$$

(E = Emotional Control, SK = Skill, and J = Judgment)