ANALYSIS OF THE PERSONAL HISTORY INVENTORY

by

L. S. Kogan
M. J. Wantman
J. W. Dunlap

A report on research conducted at the University of Rochester 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.

February 1945

CIVIL AERONAUTICS ADMINISTRATION
Division of Research
Report No. 42
Washington, D. C.

National Research Council

Committee on Selection and Training of Aircraft Pilots

Executive Subcommittee

M. S. Viteles, Chairman

E. C. Andrus
C. W. Bray
D. R. Brimhall
L. A. Carmichael
J. W. Dunlap
J. C. Flanagan
H. M. Johnson
W. E. Kellum
W. R. Miles
P. J. Rulon
G. R. Wendt

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

Dr. Dean R. Brimhall
Director of Research
Civil Aeronautics Administration
Washington 25, D. C.

February 26, 1945

Dear Dr. Brimhall:

Attached is a report entitled Analysis of the Personal History Inventory, by L. S. Kogan, M. J. Wantman, and J. W. Dunlap. This 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 being issued by the Division of Research, Civil Aeronautics Administration.

The Personal History (P-H) Inventory was originally devised as an interviewing aid for research on the aviation interview described in C.A.A. Division of Research Report No. 33. The Standard Testing Program organized through the Committee on Selection and Training of Aircraft Pilots, in 1942, furnished materials for an item analysis, for the development of scoring keys, and for further investigating the validity of this device as a predictor of success in flight training. Findings in these three areas are presented in the report.

Apart from the findings embodied in this report, the latter is an example of the extent to which the skill of numerous psychologists and the resources of many universities have been utilized through the operation of the National Research Council Committee on Selection and Training of Aircraft Pilots. The Standard Testing Program involved the voluntary cooperation of 46 psychologists throughout the country who not only contributed their services in administering tests and in gathering data but, to some extent, also made available the physical facilities of the laboratories to which they were attached at no cost to the Civil Aeronautics Administration. Without question, the research returns from funds allotted by the Civil Aeronautics Administration have been considerably extended through such contributions by individuals, universities, and other scientific agencies cooperating in the research program sponsored by the Committee on Selection and Training of Aircraft Pilots.

Cordially yours,

M S Viteles

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

MSV:ts

EDITORIAL FOREWORD

The attached report on the P-H (Personal History) Inventory grows out of the Standard Testing Program of the Committee on Selection and Training of Aircraft Pilots, National Research Council. This program, which in part grew out of plans prepared by G. R. Wendt, Wesleyan University, and J. W. Dunlap, University of Rochester, was administered at the University of Rochester by J. W. Dunlap, as Director of Research for the Committee, with the aid of M. J. Wantman and other members of the University of Rochester staff.

The Standard Testing Program, organised in 1942, involved the voluntary cooperation of 46 psychologists in various parts of the country, listed below. These psychologists tested a total of 2333 elementary and 717 secondary Civilian Pilot Training students and made available the results of examinations on a standard battery of tests for comparison with criterion data obtained from the Civil Aeronautics Administration and through the cooperation of coordinators in the C.P.T. Program.

This report of the P-H Inventory is the first of a series which will be prepared from data accumulated in the Standard Testing Program. The P-H Inventory was largely the work of J. W. Dunlap and the staff attached to the Office of the Director of Research at the University of Rochester. The latter staff was also largely responsible for the statistical treatment of data embodied in the report, which was written by L. S. Kogan, M. J. Wantman, and J. W. Dunlap. Following is the list of psychologists who cooperated in the Standard Testing Program:

Name	Institution	Location
Adams, Donald K.	Duke University	Durham, North Carolina
Allen, Clinton M.	Liberal Arts College	Oklahoma City, Oklahoma
Atkinson, Ernst A.	Montana State University	Missoula, Montana
Bathurst, James E.	Birmingham-Southern Col.	Birmingham, Alabama
Beaumont, Henry	Kentucky University	Lexington, Kentucky
Berrien, F. K.	Colgate University	Hamilton, New York
Bills, Arthur G.	University of Cincinnati	Cincinnati, Ohio
Bruce, Robert H.	University of Wyoming	Laramie, Wyoming
Caldwell, V. V.	General Extension Division, Oregon State System of Higher Education	Portland, Oregon
Crannell, Clark W.	University of Michigan	Ann Arbor, Michigan
Davis, Robert A.	University of Colorado	Boulder, Colorado
Ellson, Douglas G.	University of Mississippi	University, Mississippi
Gaskill, Harold V.	Iowa State College	Ames, Iowa
Gilmer, B. von Haller	Carnegie Inst. of Tech.	Pittsburgh, Pennsylvania
Gilliland, A. R.	Northwestern University	Evanston, Illinois
Graham, James L.	Lehigh University	Bethlehem, Pennsylvania
Grant, David A.	University of Wisconsin	Madison, Wisconsin
Hayes, George L.	University of Akron	Akron, Ohio
Hildreth, Harold M.	Syracuse University	Syracuse, New York

Name	Institution	Location
Hinckley, Elmer D.	University of Florida	Gainesville, Florida
Horton, Clark W.	Dartmouth College	Hanover, New Hampshire
Johnson, H. M.	Tulane University	New Orleans, Louisiana
Jones, Edward S.	University of Buffalo	Buffalo, New York
Kellogg, W. N.	Indiana University	Bloomington, Indiana
Koch, Adolph M.	Essex Jr. College	Newark, New Jersey
Kreezer, George L.	Cornell University	Ithaca, New York
Ligon, Ernest M.	Union College	Schenectady, New York
Lund, Max W.	University of Utah	Salt Lake City, Utah
Madden, William F.	Middlebury College	Middlebury, Vermont
Manuel, H. T.	University of Texas	Austin, Texas
Miller, Lawrence W.	University of Denver	Denver, Colorado
Miller, Vernon L.	Bowdoin College	Brunswick, Maine
Misbach, Lorenz	University of Kansas City	Kansas City, Missouri
Munn, Norman L.	Vanderbilt University	Nashville, Tennessee
Page, James D.	Temple University	Philadelphia, Penna.
Peterson, John C.	Kansas State College	Manhattan, Kansas
Sanderson, Sidney	Rutgers University	New Brunswick, N. J.
Seashore, Harold C.	Springfield College	Springfield, Mass.
Sisson, E. Donald	Louisiana State University	University, Louisiana
Terry, Paul W.	University of Alabama	University, Alabama
Tinker, Miles A.	University of Minnesota	Minneapolis, Minnesota
Willey, Clarence F.	Norwich University	Northfield, Vermont
Williams, Alexander C.	University of Maryland	College Park, Maryland
Wingfield, Robert C.	Converse College	Spartanburg, So. Car.
Wantman, Morey J.	University of Rochester	Rochester, New York
Yarborough, Joseph U.	Merit System Council, Texas Unemployment Compensation Commission	Austin, Texas

CONTENTS

SUMMARY

An analysis of the P-H Inventory initially designed as an interviewing aid, was made to investigate its possible use as a separate pencil-and-paper inventory. The inventory was administered to 1427 subjects as part of the Standard Testing Program.

The total group tested was divided into three samples for the purpose of checking the validity of scoring keys from one sample to another. Samples A and B included those with no previous flight experience, and Sample C, those with flight experience prior to testing.

From an item analysis various scoring keys were developed and the resulting inventory scores correlated with the Pass-Fail criterion and with other test scores. The following statements summarize the major findings:

1. Each of the 100 items was answered "Yes" or "No" by approximately the same proportion of cases in each sample. In this respect the samples had the same composition.

2. Item validity was found to be unstable as indicated by the variation in the correlations of the items with the Pass-Fail criterion from sample to sample and the marked variation in the number of items with P-values of .01 or less among the three samples. This instability of item validities may be dependent upon the differences in proportion of cases making up the Passer and Failer groups. Inasmuch as the number of failers is about one-tenth as large as the number of passers, its mean is much less reliable. The doubtful dependability of the criterion probably also contributes to the instability observed.

3. Keys for the P-H Inventory were developed for two types of population: (a) cases with no previous flight training (derived from Sample A and Sample B; and (b) cases with and without previous flight training (derived from all three samples). The "ultimate" key, selected for actual scoring of the inventory in related studies, includes the 27 items which obtained the same sign of ϕ coefficient in each of the three samples and P-values of .13 or less when the three samples were combined into a total group.

4. Correlations of inventory scores with the B.I., M.A.T., and M.C. were low.

5. Since the "ultimate" P-H key yields a Pass-Fail correlation of only between .30 and .40 when applied back on the samples from which it was derived, practical predictive efficiency for the P-H when applied to new samples can be expected to be low. The analysis of misclassifications of criterion prediction from P-H cutting scores similarly indicates that the predictive efficiency of the P-H keys is not satisfactory.

ANALYSIS OF THE PERSONAL HISTORY INVENTORY

INTRODUCTION

Personal histories and inventories have been employed in vocational selection to obtain data regarding the applicant's health, family, interests, cultural background, social and economic status, etc. Studies of personal data obtained from application blanks have demonstrated that biographical items can be used in predicting vocational success, praticularly on sales jobs.[1] Data obtained from studies by E. L. Kelly[2] and H. M. Johnson[3] under the auspices of the Committee on Selection and Training of Aircraft Pilots, have provided evidence that such material can also be used in predicting success or failure in flight training.

The P-H (Personal History) Inventory, discussed in this report, is essentially a device for obtaining biographical data together with self-estimates of proficiency in selected activities. The Inventory consists of 100 questions which the subject answers with "Yes" or "No" responses on a separate answer sheet. Appendix A contains sample questions and a copy of the answer sheet.

The 100 questions can be grouped into five general areas: (1) Academic Background, (2) Family and Socio-economic Background, (3) General Social Adjustment, (4) Desire to Fly, and (5) Personality as Related to Flying. Five columns on the answer sheet correspond to questions in each of the five areas, but since the subject proceeds from left to right in answering the questions in serial order he is probably not aware of the cyclic arrangement of the questions.

The P-H Inventory was originally designed for use in connection with a study of the predictive value of the interview.[4] In that study it served primarily as a convenient paper-and-pencil instrument for obtaining information from the applicant for flight training prior to the interview and as an aid to the interviewer in arriving at a judgment with respect to flight aptitude.

However, the opinion that the P-H Inventory might in itself serve as an inexpensive and convenient predictor led to the use of this questionnaire

---

[1] Viteles, M. S. Industrial psychology. New York: W. W. Norton & Co., Inc., 1932. Pp. 179-185.

[2] Kelly, E. Lowell. The relationship of background and personality factors to pilot competency. N.R.C. Division of Anthropology and Psychology, Committee on Selection and Training of Aircraft Pilots. Progress Report, September 1940.

[3] Johnson, H. M. On the actual and potential value of biographical information as a means of predicting success in aeronautical training. Washington, D. C.: C.A.A. Airman Development Division, Report No. 32, August 1944.

[4] Dunlap, Jack W., and Wantman, Morey J. An investigation of the interview as a technique for selecting aircraft pilots. Washington, D. C.: C.A.A. Airman Development Division. Report No. 33, August 1944, pp. 7-8.

in the test battery of the Standard Testing Program conducted in 1942.[5] The data obtained through the use of the P-H Inventory in the Standard Testing Program provided material for an item analysis and for the development of a scoring key used in the comparison of the P-H Inventory scores with measures of performance in flight training described in this report.[6]

PURPOSE

The aims of this analysis of the P-H Inventory were: (1) to investigate the stability of the items, i.e., to determine whether the proportion of cases answering each alternative to a given question was consistent in different samples, (2) to develop a scoring key for the P-H which would predict as efficiently as possible success in primary flight training, and (3) to study the relationship between the P-H and other tests of the standard battery.

METHODS

Complete data were obtained in 1427 cases. Of the total group, 1015 cases had had no previous flight training, and 412 cases were in primary training at the time of testing and had already had some flight experience. The "no flight hours" cases were divided into two samples. Sample A consisted of 522 cases selected at random by use of odd registration numbers. Sample B consisted of the remaining 493 cases of the "no flight hours" group. Sample C was composed of the 412 cases who had had some flight experience prior to testing.

The separation of the basic group into these three samples made it possible to construct scoring keys on the basis of item analyses on one sample and to check the validity of these keys on the other samples. The general methods used in the item analyses consisted of obtaining fourfold product-

---

[5]The Standard Testing Program was carried out at the University of Rochester in the early months of 1942 under the direction of Dr. Jack W. Dunlap, then Director of Research for the Committee on Selection and Training of Aircraft Pilots. More than 40 psychologists from various parts of the country administered a selected battery of paper-and-pencil tests to 2333 elementary students and 717 secondary students in the Civilian Pilot Training Program. This test battery consisted of: (1) the Inventory of Personal Data for Prospective Pilots (B.I.), (2) the Otis Self-Administering Test of Mental Ability, (3) the Test of Mechanical Comprehension (M.C.), (4) the Test of Aviation Information (A.I.), (5) the Personal History Inventory (P-H), and (6) the Desire-to-Fly Inventory (D-F). This program is not to be confused with the C.A.A.-National Testing Service, discussed in C.A.A. Division of Research Technical Reports Nos. 9, 19, 30, and 39.

[6]The scores derived from the item analysis were employed, **not only in** validation of the P-H Inventory on cases included in the Standard Testing Program, but also in comparing the predictive significance of scores obtained on the P-H Inventory with that of judgments made by interviewers in the study of the Aviation Interview as a predictor of success in flight training. (Dunlap, Jack W., and Wantman, Morey J. Op. cit.)

moment correlations of ϕ's[7] between answers on a particular item (Yes-No) and the criterion, passing or failing primary training, and obtaining the percentages of cases answering each item with "No." Various keys were then constructed based upon these analyses, inventory "scores" were obtained by means of these keys, and biserial correlations between the scores and the criterion data were computed.

RESULTS

<u>Proportions of Population Answering Yes-No to Items</u>. Table 1 indicates the distributions of the three samples with respect to the apportionment of Yes-No responses for the 100 questions of the Inventory. Only "No" responses were tabulated since inspection of the papers showed that very few cases skipped questions. It can be seen that these proportions range throughout the per cent scale with means at .50 and sigmas of about .30. When these item "difficulties" as based on one sample are correlated with those obtained in each of the other two samples, the correlations are .99. This high correlation indicates that each item tends to be answered "Yes" or "No" by approximately the same proportion of cases in each sample. (Although it is not shown in Table 1, it was likewise found that the proportions of "No" responses for Failers alone in Sample A ($N_f = 52$) and in Sample B ($N_f = 60$) correlated .96.) There is thus no evidence in this part of the analysis that the responses to the items are different for "independent" populations.

<u>Correlations of Items with Pass-Fail</u>. Table 2 provides the distributions of ϕ coefficients for the 100 items of the Inventory for the three samples and for the total group of 1427 subjects. It can be seen that the majority of items tend to correlate with Pass-Fail in the neighborhood of .00 and that very few items correlate as high as $\pm$.10 with the criterion in any of the samples. When all samples are thrown together, only one item correlates in excess of .10 with the criterion.

Table 3 gives the same distribution in terms of absolute correlation, i.e., without regard to sign. Here it can be seen that the mean correlation of the items with Pass-Fail ranges from .041 to .043 for the three samples, while for the total the mean correlation is .036. The number of correlations of .05 or higher in each sample is 44 and for the total population this number drops to

---

[7]Fourfold product-moment correlations between answers on each item (Yes-No) and Pass-Fail in primary flight training were obtained by calculating chi-squares for each of these comparisons and converting them into phis by dividing each chi-square by its N and taking the square root of the resulting quotient (see Kelley, T. L. <u>Statistical method</u>. New York: MacMillan Co., 1924. Ch. X, p. 259. Also Guilford, J. P. <u>Fundamental statistics in psychology and education</u>. New York: McGraw-Hill, 1942. Ch. XII, p. 246). The significance of the association was determined by evaluating the chi-squares in terms of their P-values.

TABLE 1

DISTRIBUTIONS OF 100 ITEMS IN THE P-H INVENTORY IN TERMS
OF THE PER CENT ANSWERING "NO" TO EACH ITEM

Per Cent	Sample A	Sample B	Sample C	Total
95-99	6	3	5	6
90-94	5	7	6	4
85-89	7	5	5	8
80-84	3	6	6	4
75-79	7	3	4	4
70-74	2	6	5	4
65-69	4	3	4	4
60-64	9	9	8	9
55-59	5	5	4	4
50-54	3	4	4	4
45-49	1	1	2	1
40-44	5	7	4	4
35-39	6	5	2	6
30-34	5	5	11	7
25-29	7	7	5	8
20-24	3	44	3	2
15-19	4	4	4	4
10-14	6	7	5	5
5-9	5	3	7	5
0-4	7	6	6	7
N_{items}	100	100	100	100
N_{sample}	522	493	412	1427

CORRELATIONS OF PER CENTS ANSWERING "NO"
BETWEEN SAMPLES

	Sample A	Sample B	Sample C
Sample A	—	.99	.99
Sample B		—	.99
Sample C			—
M	.50	.50	.50
σ	.30	.29	.29

N_{items} = 100

TABLE 2

DISTRIBUTION OF PHI-COEFFICIENTS FOR 100 QUESTIONS IN THE P-H INVENTORY FOR THREE SAMPLES

Ø	Sample A	Sample B	Sample C	Total
.14-.15	-	1	-	-
.12-.13	1	2	-	-
.10-.11	1	3	2	-
.08-.09	4	6	6	-
-.06-.07	8	8	8	11
.04-.05	11	7	9	10
.02-.03	10	12	11	10
.00-.01	9	12	15	15
-.02- (-.01)	15	14	20	19
-.04- (-.03)	13	14	6	20
-.06- (-.05)	13	9	13	9
-.08- (-.07)	10	5	6	5
-.10- (-.10)	3	2	3	0
-.12- (-.11)	1	2	0	0
-.14- (-.13)	0	2	1	1
-.16- (-.15)	1	0	-	-
-.18- (-.17)	-	1	-	-
N_{items}	100	100	100	100
N_{sample}	522	493	412	1427
N_f	52	60	37	149

CORRELATIONS OF PHI-COEFFICIENTS BETWEEN SAMPLES

	Sample A	Sample B	Sample C
Sample A	-	.43	.26
Sample B		-	.14
Sample C			-
$M_{Ø}$	.00	.00	-.01
$\sigma_{Ø}$	.05	.06	.05

N_{items} = 100

33 items.[8]

The correlation matrix beneath the distribution in Table 2 shows the correlations of the Ø coefficients for the items of each sample. As was to be expected from the fact that Sample A and Sample B are "no previous flight hours" subjects while the subjects in Sample C had previous flight hours, the correlation of Ø's is highest between Sample A and Sample B (r = .43). This instability of item-criterion correlations can be ascribed to the circumstance that even though the total N of each sample is approximately 500, the number of failers is only about 10% of the total and consequently the mean of the Failers is much less reliable than the mean of the Passers. Thus, all statistics such as Ø coefficients or biserial r's which are derived from the Pass-Fail dichotomy are likely to change markedly from sample to sample. Moreover, in addition to statistical considerations, the criterion of Pass-Fail is itself of dubious dependability (chiefly because of variations in standards of inspectors and manpower needs in different regions, etc.).

Table 4 gives the conversions between Ø and the P-level of significance for each sample and the total.[9] The P-level for a given Ø varies inversely with the size of the sample. Table 5 presents the distributions of P-values for the items of the Inventory in each sample. The P-values are based on the relationship between Ø and P provided in Table 4. It is to be noted that the number of items with P-values of .01 or less varies markedly among the three samples -- another indication of the instability of the correlations of the items with Pass-Fail.

<u>Construction of P-H Keys</u>. The information obtained from the item analysis described above was then used in the construction of "keys" for scoring the inventory. In order to determine how many items should be scored, several keys with varying P-levels were developed and biserial correlations obtained between the resulting "scores" and the Pass-Fail criterion.

Table 6 summarizes the Pass-Fail biserial correlations of the various keys both for the sample on which the keys were based and also on the other samples. The keys are designated in the following manner: (1) the capital letter indicates the sample upon which the item analysis was made and (2) the per cent indicates the level of P-value used in selecting the items to be scored. For example, the B (20%) key indicates a key based on the Sample B item analysis and includes the 36 items which obtained P-values of .20 or less. The "A+B" keys were based upon an item analysis in which Samples A and

---

[8]As demonstrated in Appendix B, when the Yes-No item proportion differs considerably from the criterion dichotomy proportion, the Ø obtainable, even with maximum discrimination, is lowered. In this particular study in which the criterion dichotomy was approximately 90% Passers to 10% Failers, a Ø coefficient as high as .50 is possible only when the item Yes-No dichotomy is at least .70 - .30. The distributions in Table 1 show that approximately 70 of the 100 items received less than 70% "No" responses.

[9]Obtained by finding the P-value for the chi-square equivalent of the Ø coefficient, as described in footnote 6 (supra).

TABLE 3

DISTRIBUTION OF ABSOLUTE PHI-COEFFICIENTS FOR 100 QUESTIONS IN THE P-H INVENTORY FOR THREE SAMPLES

Absolute Ø	Sample A	Sample B	Sample C	Total
.18	0	1	0	0
.17	0	0	0	0
.16	1	0	0	0
.15	0	1	0	0
.14	0	0	0	1
.13	1	3	1	0
.12	1	2	0	0
.11	1	3	0	0
.10	1	1	5	0
.09	3	4	0	1
.08	7	4	9	3
.07	9	9	9	9
.06	11	8	9	8
.05	9	8	11	11
.04	18	8	7	14
.03	7	17	7	17
.02	12	10	17	15
.01	15	14	17	10
.00	4	7	8	11
N_{items}	100	100	100	100
$M_{absolute\ Ø}$	.045	.048	.041	.036
$\sigma_{absolute\ Ø}$	.030	.037	.030	.025
N_{sample}	522	493	412	1427

TABLE 4

CONVERSION BETWEEN Ø AND P-VALUE

Ø	Sample A P-value	Sample B P-value	Sample C P-value	Total P-value
±.13 and over	Less than .01	Less than .01	.01 or less	Less than .01
±.12	.01	.01	.01	Less than .01
±.11	.01	.01	.02	Less than .01
±.10	.02	.03	.04	Less than .01
±.09	.04	.05	.06	Less than .01
±.08	.07	.08	.10	Less than .01
±.07	.11	.12	.16	.01
±.06	.17	.18	.22	.02
±.05	.25	.27	.31	.06
±.04	.36	.37	.42	.13
±.03	.49	.51	.54	.26
±.02	.65	.66	.68	.45
±.01	.82	.82	.84	.70
±.00	1.00	1.00	1.00	1.00
N	522	493	412	1427

B were combined into a "no flight hours" group. The "A:B (same sign items, 50%)" key includes those items which obtained Ø coefficients with the same sign and P-values of .50 or less in both Sample A and Sample B separately.

As is to be expected, the keys correlated most highly with the Pass-Fail criterion when applied to the sample upon which they are based. There is considerable loss in predictive efficiency when the keys are applied to the other samples, even between Samples A and B which were randomly selected from a group homogeneous with respect to lack of previous flight experience. Apparently the most efficient of the keys based on a single sample is the A (20%) key which correlates .536 in Sample A, .343 in Sample B, and .210 in Sample C.

Selection of the "Ultimate" Key. In the selection of items for a key to be used for obtaining a P-H score (for possible field use) several requirements were set up: (1) the sign of the item Ø coefficient should be the same for each of the three samples; (2) the item P-values should be low for each of the three samples; and (3) the number of scored items should be as large as possible.

The second requirement, however, could not be met. As shown in Table 7, even when items with P-values as high as .50 were included (Key I), the total number of available items was only 13. Five other possible keys, based on P-values in the total samples item analysis, were then constructed. Of these, Key VI was considered to be the best and was selected as the "ultimate" key to be used in actual scoring of the inventories. This "ultimate" key includes

TABLE 5

DISTRIBUTIONS OF P-VALUES FOR 100 QUESTIONS
IN THE P-H INVENTORY FOR THREE SAMPLES

P	Sample A	Sample B	Sample C	Total
Less than .01	2	5	0	2
.01	2	5	1	12
.02-.04	4	1	5	4
.05-.09	7	8	0	12
.10-.14	9	9	9	12
.15-.19	11	8	9	0
.20-.24	0	0	9	0
.25-.29	9	8	0	21
.30-.34	0	0	11	0
.35-.39	18	8	0	0
.40-.44	0	0	7	0
.45-.49	7	0	0	13
.50-.54	0	17	7	0
.55-.59	0	0	0	0
.60-.64	0	0	0	0
.65-.69	12	10	17	0
.70-.74	0	0	0	14
.75-.79	0	0	0	0
.80-.84	15	14	17	0
.85-.89	0	0	0	0
.90-.94	0	0	0	0
.95-.99	0	0	0	0
1.00	4	7	8	10
N_{items}	100	100	100	100
N_{sample}	522	493	412	1427

TABLE 6

VALIDITY COEFFICIENTS* FOR SAMPLES BY SELECTED KEYS

Key	N_{items}	Sample A	Sample B	Sample C
A (1%)	4	.330	.123	.130
A (5%)	8	.419	.109	.058
A (10%)	15	.451	.157	.171
A (20%)	35	.536	.343	.210
A (35%)	96	.442	.388	.203
B (1%)	10	.183	.565	.071
B (5%)	15	.184	.621	.053
B (10%)	19	.218	.655	.055
B (20%)	36	.152	.672	.046
B (35%)	93	.191	.654	.166
A+B (1%)	10	.245	.473	.067
A+B (5%)	45	.471	.439	.129
A:B (same sign)	64	.400	.490	.201
A:B (same sign, 50%)	63	.317	.567	.171
A:B:C (same sign, 13%)	27	.295	.344	.404
N_{sample}		512	493	412
N_f		52	60	37

* Correlations are biserial with respect to Pass-Fail

the 27 items which obtained the same sign of ∅ in each of the three independent samples and P-values of .13 or lower in the total sample. It was this key which was used in scoring the P-H answer sheets in the Aviation Interview[10] study and in the further analysis described below.

Distribution of ∅ Coefficients of the "Ultimate" P-H Key. The distribution of absolute ∅ coefficients for the 27 items of the "ultimate" key is given in Table 8. The mean ∅ for each sample is: .053 for Sample A, .063 for Sample B, and .055 for Sample C. Here it can be seen again that the highest ∅ obtained is only .18. Likewise it is obvious that some of these "best" items go as low as .01 in some of the three samples. Beneath the distribution of ∅'s are the mean scores of Passers and Failers as obtained by applying the "ultimate" key back on the samples; the corresponding biserial correlations with Pass-Fail for each sample are also indicated.

[10] Dunlap, Jack W., and Wantman, Morey J. Op. cit.

TABLE 7

POSSIBLE TYPES OF P-H KEYS BASED ON THREE INDEPENDENT SAMPLES

		Number of Items Available
I.	(1) Same sign of Ø (each sample) (2) Approximate P of .50 or less in each sample	13
II.	(1) Same sign of Ø (each sample) (2) P of .50 or less in total sample combined	35
III.	(1) Same sign of Ø (each sample) (2) P of .26 or less in total sample combined	32
IV.	P of .13 or less in total sample combined	42
V.	P of .06 or less in total sample combined	29
*VI.	(1) Same sign of Ø (each sample) (2) P of .13 or less in total sample combined	27

*Key VI was chosen as the "ultimate" key.

Intercorrelations of P-H Scores with B.I., M.A.T., M.C., and Pass-Fail. Table 9 presents the intercorrelation of scores based on keys from Samples A and B, the "ultimate" key based on all three samples, and the B.I., M.A.T., and M.C. test scores. The number of subjects shows a decrease from the original samples because not all of the scores on the latter tests were available. Inspection of the matrices for the three samples shows that all of the P-H keys are fairly highly intercorrelated. The key yielding the highest Pass-Fail correlation varies from sample to sample. The best key for Sample A is "A+B (5%)," for Sample B it is "A:B (same sign, 50%)." Considering just Sample A and Sample B, which are the subjects with no previous flight hours, perhaps the best key is "A+B (5%)." From the standpoint of efficiency in all three samples, the best key is obviously the "ultimate" key based on all three samples.

Correlations of the other tests with the P-H scores are fairly low with the B.I. yielding the highest intercorrelation. The correlation of the B.I. with the "ultimate" P-H key is .255 in Sample A, .213 in Sample B, and .212 in Sample C. The M.A.T. shows no correlation with the "ultimate" P-H key while the M.C. shows a very low positive correlation.

In view of the fact that the "ultimate" P-H key was derived in part from each of the samples considered, it is not surprising that correlations of this key with the Pass-Fail criterion should exceed the correlations of the other three major tests used (B.I., M.A.T., M.C.) with the same criterion. It is, however, more noteworthy that, even under these conditions, in Sample

TABLE 8

DISTRIBUTION OF ABSOLUTE PHI-COEFFICIENTS FOR 27 ITEMS OF "ULTIMATE" KEY OF P-H INVENTORY FOR THREE SAMPLES

Absolute Ø	Sample A	Sample B	Sample C
.18	-	1	1
.17	-	-	-
.16	1	-	-
.15	-	-	-
.14	-	-	-
.13	-	1	-
.12	-	1	-
.11	1	1	-
.10	-	1	1
.09	-	1	-
.08	4	2	6
.07	2	2	3
.06	4	2	3
.05	5	5	4
.04	1	4	2
.03	3	4	-
.02	2	1	4
.01	4	1	3
N_{items}	27	27	27
$M_{absolute\ Ø}$	.053	.063	.055
$\sigma_{absolute\ Ø}$	.033	.038	.029
N_{sample}	522	493	412
M_p	19.59	19.39	19.81
$M_{\bar{p}}$	18.14	17.63	17.86
σ_T	2.55	2.70	2.44
Bis. r	.295	.344	.404

TABLE [illegible]

VALIDITY COEFFICIENTS OF SELECTED P-H KEYS AND INTERCORRELATIONS

Sample A

	1	2	3	4	5	6	7	8	9
1. Pass-Fail (biserial)	-	.268	.490	.415	.335	.182	.186	.318	.298
2. P-H Key A+B (1%)		-	.586	.526	.490	.086	.039	.130	.445
3. P-H Key A+B (5%)			-	.825	.792	.170	.088	.161	.673
4. P-H Key A:B (same sign)				-	.947	.242	.051	.173	.771
5. P-H Key A:B (same sign, 50%)					-	.263	.081	.168	.760
6. B.I. (+1% A)						-	.144	.226	.255
7. M.A.T.							-	.434	.045
8. M.C.								-	.153
9. P-H Ultimate Key, A:B:C (same sign, 13%)									-
M	.89	6.56	19.33	22.50	42.17	12.48	53.54	52.13	19.45
σ	1.00	2.76	7.12	8.16	4.33	2.84	9.72	7.14	2.56

N = 501
N_p = 447

Sample B

	1	2	3	4	5	6	7	8	9
1. Pass-Fail (biserial)	-	.490	.423	.467	.534	.254	.218	.272	.324
2. P-H Key A+B (1%)		-	.642	.613	.603	.189	.094	.216	.494
3. P-H Key A+B (5%)			-	.828	.826	.145	.068	.162	.641
4. P-H Key A:B (same sign)				-	.950	.139	-.004	.186	.733
5. P-H Key A:B (same sign, 50%)					-	.197	.014	.172	.759
6. B.I. (+1% A)						-	.168	.332	.213
7. M.A.T.							-	.441	.007
8. M.C.								-	.102
9. P-H Ultimate Key, A:B:C (same sign, 13%)									-
M	.88	6.26	18.72	21.65	41.73	12.31	52.99	51.45	19.22
σ	1.00	2.55	7.03	8.78	4.51	2.97	9.94	7.66	2.72

N = 473
N_p = 413

Sample C

	1	2	3	4	5	6	7	8	9
1. Pass-Fail (biserial)	-	.014	.143	.217	.202	.188	.073	.304	.401
2. P-H Key A+B (1%)		-	.524	.531	.541	.065	-.101	.098	.393
3. P-H Key A+B (5%)			-	[illegible]	.797	.162	.012	.213	.607
4. P-H Key A:B (same sign)				-	.958	.235	-.054	.213	.699
5. P-H Key A:B (same sign, 50%)					-	.243	-.028	.221	.708
6. B.I. (+1% A)						-	.084	.335	.212
7. M.A.T.							-	.433	.054
8. M.C.								-	.187
9. P-H Ultimate Key, A:B:C (same sign, 13%)									-
M	.91	6.68	19.51	23.17	42.18	13.44	51.61	51.19	19.65
σ	1.00	2.16	6.20	7.30	3.74	2.87	9.58	7.11	2.45

N = 410
N_p = 375

A, the correlation of the M.C. with the Pass-Fail criterion exceeds that of the "ultimate" P-H key. The M.C. consistently showed the highest correlations with Pass-Fail of the three major tests used in the Standard Testing Program.

It is interesting to note how the B.I. biserial correlations with Pass-Fail compare with P-H biserial correlations. These comparisons are made below for selected keys using data from Table 6. The underlined correlations are those which were obtained in samples not used in the construction of the particular key being considered. For the underlined values, which represent the unbiased[11] correlations between the P-H and the criterion, errors of the correlations are all within range of sampling errors of the correlations for the B.I.

	Sample A	Sample B	Sample C
P-H Key A (20%)	.54	.34	.21
P-H Key B (20%)	.15	.67	.05
P-H Key A + B (5%)	.47	.44	.13
P-H Key A:B (same sign)	.40	.49	.20
P-H Key A:B:C (same sign, 13%)	.30	.34	.40
P-H Key B.I.	.18	.25	.19

Misclassifications or Errors of Criterion Prediction from P-H Cutting Scores. The two keys constructed for the P-H Inventory which seem to have the most logical applicability are: (1) The key based on Sample A and B (no flight hour groups) which is comprised of items maintaining the same sign in both samples (Key A:B, same sign, 50%) and (2) the "ultimate" key based on all three samples (Key A:B:C, same sign, 13%).

In order to compare the efficiency of these two keys as applied to the three samples in the study with the maximum efficiency possible and also with the efficiency of a key based on chance, the procedure described below was followed.[12]

Fourfold charts were set up in such a way that the distribution of P-H scores obtained by scoring with each key was cut at a point which dichotomized the distribution in as equivalent a manner as possible to the dichotomy imposed by the Pass-Fail criterion. For example, if the percentage of Passers in the group was 90% the distribution of P-H scores was cut so that as close to 90% of the cases were above the cutting score as was permissible from the

---

[11] I.e., those correlations obtained when the key was applied to samples other than the sample upon which the key was developed.

[12] This procedure was suggested by Dr. H. M. Johnson, Professor of Psychology, Tulane University.

score groupings. A fourfold chart resulted such as is seen in the diagram below:

	Above Cutting Score	Below Cutting Score	Total	
Pass	α	β	90%	p_2
Fail	γ	δ	10%	q_2
	90%	10%	100%	
	p_1	q_1		

Here we see that 90% of the subjects passed and 90% of the subjects were above the cutting score on the Inventory. It is obvious that the subjects falling in categories α and δ are correctly classified by the Inventory, i.e., α are the passers who are above the cutting score and δ are the failers who are below the cutting score. Categories β and γ, in contrast, are the subjects who have been misclassified. If the Inventory had discriminated perfectly we would expect zero entries in categories β and γ. Our expectations in the cells according to "chance" would be based on the marginal totals. Thus, since β is p_2q_1 and γ is p_1q_2, the number of chance misclassifications to be expected from a non-discriminating equivalent dichotomy of inventory and criterion would be the sum of p_2q_1 and p_1q_2 or, in the diagram above, we would expect 18% of the subjects to be misclassified. An inventory with some degree of discrimination will hence misclassify less than 18% of the subjects or in the case of perfect discrimination, zero per cent.

The number of misclassifications for each sample on the P-H Inventory to be expected by chance, the number actually obtained, and minimum possible number are presented in Table 10.

It is of interest to note that the cutting scores for the three samples turned out to be so similar, even though determined separately for each sample. The actual distributions of scores are presented in Appendix G.

By comparing the entries in the "By Chance" column with the entries in the "By P-H" column we can tell how much better than chance the P-H has predicted Pass-Fail. The entries in the column under "Minimum" are not zero because the distribution of scores could not be cut in exactly the same proportion as the criterion dichotomy. From these tables it is obvious that the efficiency of the P-H keys is only slightly better than one would expect from chance alone.[13] Dichotomizing the inventory scores, however, partially con-

---

[13]The standard error of this statistic was not derived and it may well be that the proportions obtained were within the range of sampling error. However, the fact that all of the "shifts" were towards improved classification probably renders the null hypothesis untenable in this situation.

TABLE 10

P-H INVENTORY - MISCLASSIFICATIONS UNDER CONDITIONS OF MOST EQUIVALENT DICHOTOMY BETWEEN INVENTORY AND CRITERION

Key A:B (same sign, 50%)		Number of Misclassifications*		
Sample	Cutting Score	By Chance	By P-H	Minimum
Sample A N = 501 N_f = 54	36	95 (19%)	71 (15%)	3 (1%)
Sample B N = 473 N_f = 55	36	99 (21%)	71 (15%)	3 (1%)
Sample C N = 410 N_f = 35	36	60 (15%)	52 (13%)	4 (1%)

Key A:B:C (same sign, 13%)		Number of Misclassifications		
Sample	Cutting Score	By Chance	By P-H	Minimum
Sample A N = 501 N_f = 54	15	87 (17%)	73 (15%)	13 (3%)
Sample B N = 473 N_f = 55	15	87 (18%)	75 (16%)	13 (3%)
Sample C N = 410 N_f = 35	16	73 (18%)	53 (13%)	11 (3%)

* The per cents indicate the proportion of the total N of the sample.

ceals the extent to which the inventory is related to the criterion, throughout the total range of scores. This may account for the fact that although the per cent misclassified by the P-H is quite similar from sample to sample, the biserial correlations (as shown in Table 9) between Key A:B:C (same sign, 13%) and Pass-Fail range from .298 to .401.

SUMMARY

The P-H Inventory was administered to 1427 subjects as part of the Standard Testing Program. The total group tested was divided into three samples and an item analysis made of the responses in order to check item validity from sample to sample, and to develop scoring keys from which inventory scores could be obtained and compared with the Pass-Fail criterion and other test measures. The resulting findings were:

1. Although the items were answered "Yes" or "No" by approximately the same number of subjects in each sample, correlations of the items with the criterion varied markedly from sample to sample, probably as a result of the small number of failers and the doubtful dependability of the criterion.

2. The ultimate key selected for scoring of the inventory contained 27 items which maintained the same sign of ϕ in each of the three groups and P-values of .13 or less for the total group.

3. Intercorrelations of inventory scores with BI., M.A.T., and M.C. were low.

4. Since the ultimate P-H key produces a Pass-Fail correlation of only between .30 and .40 when applied back on the samples from which it was derived, the probability of practical predictive efficiency for the P-H when applied to new samples is low.

5. Misclassifications of criterion prediction from P-H cutting scores are only slightly less than are to be expected by chance.

APPENDIX A

SAMPLE QUESTIONS FROM P-H INVENTORY
and
ANSWER SHEET FOR P-H INVENTORY

SAMPLE QUESTIONS FROM
PERSONAL HISTORY INVENTORY

1. Were any of the following your favorate subjects in school -- English, History, Music, Art, Languages?
2. Were both of your parents living while you were in high school?
3. Did you have many fights with other boys in school?
4. Have you been interested in flying for more than three years?
5. Did your parents approve of your enrolling in aviation?

81. Are you a good diver from a springboard?
82. Did you earn part of your way through high school to college?
83. Do you think "petting" is wrong?

Answer Sheet for P-H INVENTORY

Name .... Age .... Single .... Married ....
Print Last First Initial

Schooling 9, 10, 11, 12, 13, 14, 15, 16, 17 (circle highest grade) Name of College ....

Major Course in College .... Location of College ....

	Yes	No		Yes	No		Yes	No		Yes	No		Yes	No
.			2.			3.			4.			5.		
5.			7.			8.			9.			10.		
.			12.			13.			14.			15.		
5.			17.			18.			19.			20.		
1.			22.			23.			24.			25.		
5.			27.			28.			29.			30.		
1.			32.			33.			34.			35.		
5.			37.			38.			39.			40.		
1.			42.			43.			44.			45.		
5.			47.			48.			49.			50.		
1.			52.			53.			54.			55.		
5.			57.			58.			59.			60.		
1.			62.			63.			64.			65.		
5.			67.			68.			69.			70.		
1.			72.			73.			74.			75.		
5.			77.			78.			79.			80.		
1.			82.			83.			84.			85.		
5.			87.			88.			89.			90.		
1.			92.			93.			94.			95.		
5.			97.			98.			99.			100.		

Sport	Attitude*			Did you go out for this sport in		Did you make the varsity in	
	L	I	D	H. S.	Col.*	H. S.	Col.*
Football							
Basketball							
Baseball							
Soccer							
Rowing							
Track							
Swimming							
Tennis							
Boxing							
Wrestling							

* L = like
I = indifferent
D = dislike

* H. S. = high school
Col. = college

DIRECTIONS FOR USE OF THE ANSWER SHEET WITH THE P-H INVENTORY

There are 100 questions in the P-H Inventory booklet. Answer each question by placing a check mark (√) on the line under the appropriate heading of "yes" or "no" next to the number that corresponds to the number of the question. Note that the question numbers on the answer sheet are arranged *across the page.* At the beginning of each new line of the answer sheet, be sure to check the number of the question in the booklet with the number on the answer sheet. *Do not omit any question.*

When you have finished with the 100 questions in the booklet fill in the information regarding sports at the bottom of the answer sheet by placing check marks in the proper columns. In the first column marked "Attitude," indicate your attitude towards each sport by checking whether you like, dislike, or are indifferent to that sport. In the next two columns indicate your participation in these sports. You may have no marks, one mark, or up to four marks in these last two columns for each sport.

EXAMPLE: (Check marks show how one man filled in this part of the blank)

Sport	Attitude*			Did you go out for this sport in		Did you make the varsity in	
	L	I	D	H. S.	Col.*	H. S.	Col.*
Football	√			√	√	√	
Basketball	√			√	√		√
Baseball	√			√			
Soccer			√				
Rowing		√					
Track	√				√		

* L = like
I = indifferent
D = dislike

* H. S. = high school
Col. = college

This man liked football, went out for it in high school and college, but made the varsity only in high school. He liked baseball, went out for it in high school, but did not make the varsity. He liked track, went out for it in college, but did not make the varsity.

APPENDIX B

THE MAXIMALITY OF ITEM ϕ COEFFICIENTS WHEN THE CRITERION DICHOTOMY IS SUCH THAT 90% OF THE CASES ARE PASSERS

APPENDIX B

THE MAXIMALITY OF ITEM Ø COEFFICIENTS WHEN THE CRITERION DICHOTOMY IS SUCH THAT 90% OF THE CASES ARE PASSERS

The following note describes certain limitations of the Ø coefficient computed from a two-by-two chart when the criterion dichotomy and the item dichotomy proportions differ. As shown below, even if items are operating with their best possible efficiency the size of the maximum Ø coefficient is considerably reduced from 1.00 when, as in the present instance of a 90%-10% criterion dichotomy, the item dichotomy proportions deviate from 90%-10%.

In computing the Ø coefficients the two-by-two chart is set up as follows:

	No	Yes	Total	
Pass	α	β	.90	p_2
Fail	γ	δ	.10	q_2
	p_1	q_1	1.00	

Here α = per cent of passers checking "No"

β = per cent of passers checking "Yes"

γ = per cent of failers checking "No"

δ = per cent of failers checking "Yes"

The formula for $\phi = \dfrac{\alpha\delta - \beta\gamma}{\sqrt{p \times q \times .90 \times .10}}$

Because of the nature of the Ø coefficient, a Ø of 1.00 can be attained in the above case only when $\alpha = p_1 = .90$. As p_1 deviates from .90 the maximum Ø which could be reached under the optimum conditions of discrimination is indicated in the accompanying graph. For example, when the criterion Pass-Fail ratio is $\frac{.90}{.10}$ and the item is dichotomised with 50% choosing each alternative, the maximum Ø attainable can be seen to be .33.

The two-by-two chart under these conditions of best discrimination for an item p_1 of .50 is as follows:

	No	Yes	Total	
Pass	.50	.40	.90	p_2
Fail	.00	.10	.10	q_2
	.50	.50	1.00	
	p_1	q_1		

In this situation because of the .50 - .50 point distribution of the item, the maximum number of passers who can be classified correctly by the item is only 50% of the total cases even though the failers (10%) are all correctly classified. Hence, passers constituting 40% of the total must of necessity be misclassified and consequently the ceiling of the ϕ is .33.

Other maximum ϕ's at varying p-levels of the items as associated with a criterion p_2 of .90 are indicated in the graph.[14]

The efficiency of discrimination indicated by a ϕ is therefore influenced by the specific proportion of subjects associated both with the criterion dichotomy and the item dichotomy. The greater the disparity between the positions of the points of truncation of the two variables the less likelihood there is of a ϕ approaching 1.00. This gradient away from a maximum ϕ of 1.00 as $p_1 \neq p_2$ is not symmetrical, however, except when the p_2 of the criterion is .50.

In the case where $p_2 = .90$ it can thus be seen that a ϕ of .50 or over is possible only when the item p_1 associated with passing on the criterion is in excess of approximately .70. With many items having a p_1 of less than .50 we must expect many low ϕ's even under potentially favorable conditions of discrimination.

Since the ϕ is related to X^2 such that $X^2 = N\phi^2$, it can be seen that X^2 and hence the P-level of significance are affected only by ϕ and N and are thus subject to the same limitations of criterion and item disparity of dichotomization as ϕ.

The main conclusion to be kept in mind is that item dichotomies must be as equivalent as possible to the criterion dichotomy if efficient discrimination is desired.

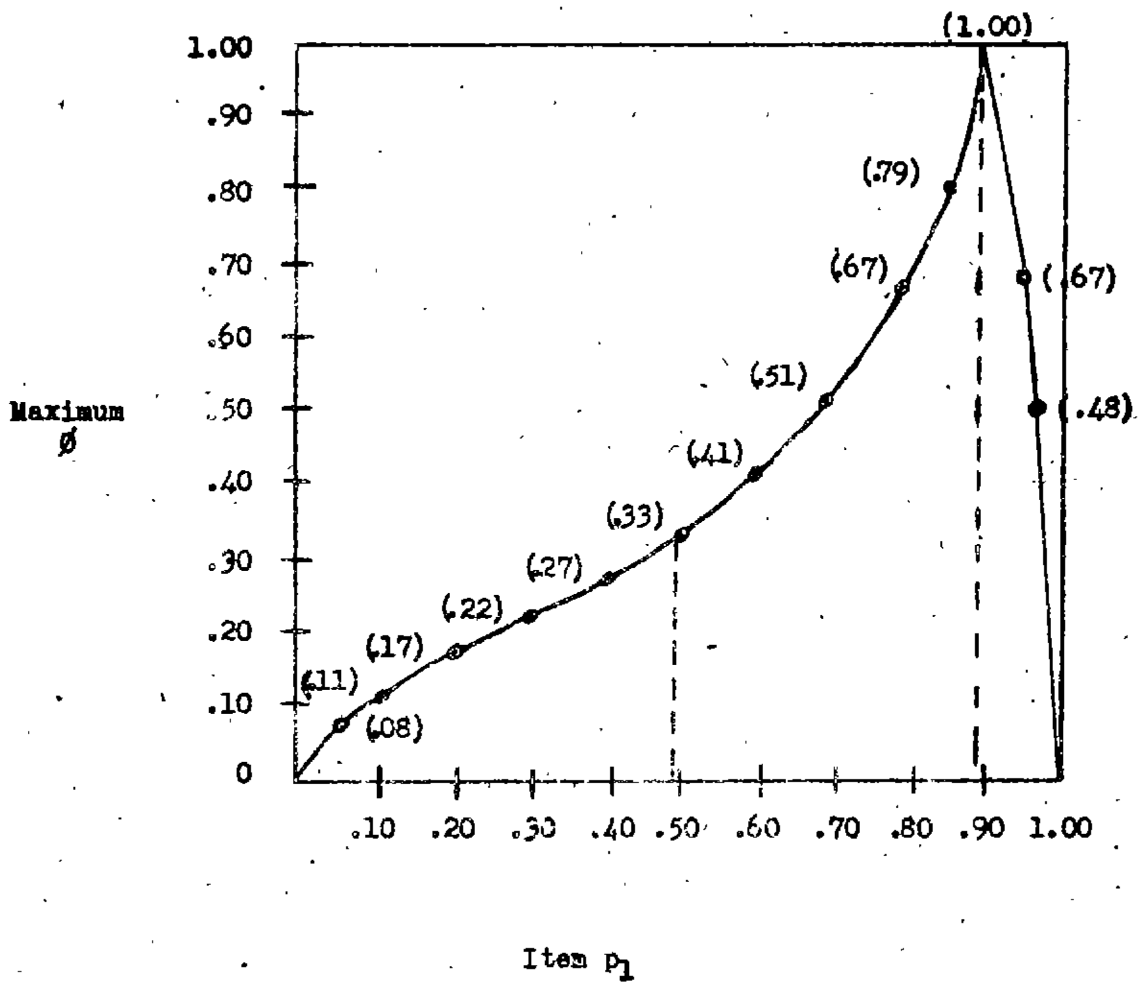
---

[14]<u>Editor's Note</u>. This graph represents the maximum phi obtainable for a particular criterion dichotomy, namely, 90% passers and 10% failers. Professor H. M. Johnson has shown that the maximum phi can be determined for any combination of criterion and item dichotomy proportions from a 2 x 2 distribution by means of the following formula derived for the general case. This formula is as follows:

$r = \Delta/N \sqrt{p_1 q_2} \cdot \sqrt{p_2 q_1}$, in which p_1 is the proportion of test passers and p_2 the proportion of course passers in the sample of N individuals, while $q_1 = 1-p_1$, $q_2 = 1-p_2$, $\Delta = (AB) - (A)(B)/N$. Then $(\Delta/N)_{max} = p_1 q_2$ or $p_2 q_1$, whichever is the smaller, and substituting in the equation,

$r_{max} = \sqrt{p_1 q_2} / \sqrt{p_2 q_1}$ or $r_{max} = \sqrt{p_2 q_1} / \sqrt{p_1 q_2}$. For a discussion of the development of the formula given above, see: Johnson, H. M. General rules for predicting the selectivity of a test. <u>Amer. J. Psychol.</u>, 1942, 55, 436-442.

MAXIMUM Ø WHEN CRITERION DICHOTOMY IS
90% PASSERS AND 10% FAILERS



APPENDIX C

DISTRIBUTIONS OF INVENTORY SCORES FOR SAMPLES A, B, AND C

APPENDIX C

DISTRIBUTIONS OF INVENTORY SCORES FOR SAMPLES A, B, & C

P-H KEY [illegible] (Same sign, 50%)

Sample A				Sample B				Sample C			
Score	Total Group	Passers	Failers	Score	Total Group	Passers	Failers	Score	Total Group	Passers	Failers
53	1		1	51	5	5		52	2	2	
52	2	2	0	50	4	4		51	2	2	
51	3	3	0	49	10	10		50	4	4	
50	12	12	0	48	23	23		49	5	5	
49	12	12	0	47	28	28		48	17	16	1
48	25	25	0	46	31	30	1	47	19	18	1
47	23	21	2	45	31	28	3	46	30	29	1
46	39	36	3	44	44	43	1	45	32	28	4
45	32	32	0	43	47	46	1	44	37	36	1
44	55	50	5	42	49	45	4	43	55	48	7
43	43	37	6	41	34	32	2	42	46	42	4
42	41	37	4	40	28	21	7	41	31	28	3
41	43	40	3	39	24	19	5	40	32	32	0
40	42	38	4	38	29	26	3	39	30	29	1
39	25	22	3	37	28	21	7	38	22	20	2
38	23	20	3	36	19	13	6	37	15	12	3
37	29	26	3	35	10	7	3	36	13	10	3
36	14	10	4	34	8	3	5	35	9	7	2
35	14	9	5	33	5	5	0	34	3	2	1
34	12	8	4	32	4	2	2	33	2	1	1
33	3	2	1	31	4	3	1	32	3	3	0
32	4	3	1	30	5	1	4	31	1	1	0
31	1	1	0	27	1	1	0				
30	3	1	2	26	1	1	0				
				25	1	1	0				
	501	447	54		473	418	55		410	375	35

APPENDIX C

DISTRIBUTIONS OF INVENTORY SCORES FOR SAMPLES A, B, & C

P-H KEY A:B:C (same sign, 13%)

Sample A

Score	Total Group	Passers	Failers
26	1	1	
25	5	5	
24	7	6	1
23	42	40	2
22	57	54	3
21	77	70	7
20	78	71	7
19	65	61	4
18	61	53	8
17	33	28	5
16	34	28	6
15	23	19	4
14	11	6	5
13	5	4	1
12	2	1	1
	501	447	54

Sample B

Score	Total Group	Passers	Failers
26	1	1	
25	3	3	
24	10	10	
23	30	29	1
22	48	46	2
21	81	76	5
20	62	56	6
19	67	59	8
18	55	47	8
17	37	26	11
16	37	34	3
15	21	15	6
14	9	6	3
13	5	5	0
11	1	1	0
10	6	4	2
	473	418	55

Sample C

Score	Total Group	Passers	Failers
26	1	1	
25	5	5	
24	18	18	
23	22	21	1
22	47	47	0
21	55	53	2
20	85	75	10
19	50	46	4
18	50	47	3
17	31	30	1
16	25	17	8
15	12	8	4
14	7	6	1
13	1	1	0
11	1	0	1
	410	375	35