TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FAA-AM-71-35		

4. Title and Subtitle	5. Report Date
PERSONALITY ASSESSMENT IN AVIATION: AN ANALYSIS OF THE ITEM AMBIGUITY CHARACTERISTICS OF THE 16PF AND MMPI	July 1971
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Roger C. Smith, Ph.D.	
9. Performing Organization Name and Address	10. Work Unit No.
FAA Civil Aeromedical Institute P. O. Box 25082 Oklahoma City, Oklahoma 73125	11. Contract or Grant No.
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address	
Office of Aviation Medicine Federal Aviation Administration 800 Independence Avenue, S.W. Washington, D. C. 20590	OAM Report
	14. Sponsoring Agency Code

15. Supplementary Notes

This research was conducted under Tasks AM-A-70-PSY-20 and AM-A-71-PSY-20.

16. Abstract

Devices such as the 16PF and MMPI have been widely employed in the evaluation of personnel in aviation settings. The present study investigated the problem of item ambiguity (the degree to which an item elicits multiple interpretation) which may limit the utility of such devices when used in screening procedures. Subjects completed either the 16PF or the MMPI while concurrently rating each item on a five-point ambiguity scale. The ambiguity for each item was determined and the relationship between ambiguity and sex of the respondent, the individual factor scales, and the scores of subjects on the scales was considered. The implications of the findings for item construction and use of the test in various applications were discussed.

17. Key Words	18. Distribution Statement
Personality Assessment Psychological Evaluation	Availability is unlimited. Document may be released to the National Technical Information Service, Springfield, Virginia 22151, for sale to the public.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	6	$3.00

Form DOT F 1700.7 (8-69)

PERSONALITY ASSESSMENT IN AVIATION: AN ANALYSIS OF THE ITEM AMBIGUITY CHARACTERISTICS OF THE 16PF AND MMPI

I. Introduction.

Devices such as the 16PF and MMPI have been widely employed in the evaluation of personnel in aviation settings. The 16PF in particular has been used in the screening of air traffic controllers for adjustment problems which might influence their ability to function as controllers.[7] Such inventories have proven to be useful in such assessments, but could still be improved with respect to utility and validity.

A problem with the inventories which respondents often mention is that it is sometimes difficult to answer items because of uncertainty about the intended meaning of the item. In other words, an item may elicit a number of different interpretations, any of which could be chosen as the basis for the respondent's answer. While this may only be an annoyance for the respondent, it is of considerable importance to the user of the test, since increased ambiguity in items may have a significant effect on the validity of the technique. Anastasi[1] has argued that if an item has several possible interpretations, there is a greater probability of random, or unwanted, variance in responses than if an item only has one or two possible meanings. On the other hand Broen[3] has presented an argument suggesting that ambiguity may, under some circumstances, enhance validity. It would, therefore, seem advisable to determine the item ambiguity characteristics of assessment devices as an adjunct to other item analysis procedures.

To date, the empirical description of item ambiguity has focused solely on the MMPI. Harris and Baxter[5] had subjects rate the degree of ambiguity in each MMPI item and found that, while item ambiguity was relatively low for most items, there were variations in ambiguity level associated with sex, scales, and personality characteristics of the respondents. While there have been no reported studies of the ambiguity variable for the 16 PF, it has been noted by Harsh[6] that, "Many of the items have an objectionable ambiguity which was criticized in older questionnaires, thus arousing a rather uncooperative attitude in college students. . . ." Thus, it was the primary purpose of the present study to extend the approach taken by Harris and Baxter[5] to the 16PF. In addition to determining the item-ambiguity characteristics of the A and B forms of the 1967–68 revision of the 16PF, it was also decided to assess the consistency of item-ambiguity characteristics by replicating the Harris and Baxter[5] MMPI ambiguity ratings. The replication will also serve the purpose of establishing a basis for comparison of the general item ambiguity characteristics of the 16PF and MMPI.

The assistance of Sharon Bowles, Karen Freeman, Jennifer Jordan, Terry Lazar, and Dennis Waggoner in acquisition and analysis of the data is gratefully acknowledged.

II. Method.

A. *Subjects.* The subjects for this study were paid student volunteers from a large southwestern university. All subjects were recruited by a campus research support agency to participate in a learning experiment. Participation in the ambiguity study was coincident with the learning study, but was not a factor in the selection of subjects. The total of 227 volunteers included 144 males and 83 females.

B. *Procedure.* The subjects were alternately designated to take either the 16PF or the MMPI. The A and the B forms of the 1967–68 16PF revision were administered to equal numbers of subjects receiving the 16PF.

Before the administration of the inventory, *E* presented the standard inventory instructions to the subject. At the completion of the standard instructions, the procedure to be followed for rating ambiguity was indicated. The instructions for this rating followed those used by

Harris and Baxter[5] and the five-point scale devised by these authors was employed. The following instructions were presented:

> The second task is to rate each item in the inventory as you go along according to how ambiguous, that is how indefinite, vague, or unclear, you find the item. How much difficulty do you have in responding to the item for one, or all of several reasons—the specific wording of the statement, the vagueness of it, the many alternative meanings some items might have for you, or the generality of the statement with regard to time, place, or circumstances? Make your ratings in the following way: note the card taped to the front of your desk; it has a five point scale typed upon it. If an item is "not at all ambiguous, very easy to answer" score it a one. If the item is slightly ambiguous to you, code it a two; if moderately ambiguous, code it a three; if very ambiguous, a four; and finally, if extremely ambiguous, mark it a five. Do not shy away from using the highest or lowest numbers if you feel these apply to any item.
>
> Remember that we are interested only in the degree of ambiguity of the statement as a cause of difficulty in answering the item. In the rating we are not at all interested in the truth or falsity of the item—that is taken care of in your initial response.

After scoring the profiles from both inventories it was determined that 105 MMPI, 53 16PF-A, and 52 16PF-B records were completed with respect to both answers and ambiguity ratings.

III. Results.

As in the Harris and Baxter[5] study, the percentage of subjects who assigned a rating of "two" or more, i.e., rated an item at least "slightly ambiguous," correlated approximately .94 with the mean ambiguity rating for each item on both the 16PF and the MMPI. Since use of the percentage value, hereafter referred to as the item ambiguity value, tended to reduce the effects of the marked skewness of the ratings for many items, it was used as the measure of ambiguity for all subsequent analyses but those concerned with the tendency of individual subjects to rate items as ambiguous (subject ambiguity).

A. *16PF*. The mean item ambiguity value for Form A was .41 (S.D. = .23), as 23% of the items were considered ambiguous by at least one-half the subjects taking this form. For the B form, the means was .40 (S.D. = .22) and 19.8% of the items were judged ambiguous by a majority of the subjects. There were no significant differences between forms for these values.

There was a trend for ambiguity ratings to vary as a function of sequence in both forms of the 16PF; Form A, F (3,366) = 5.30, $p < .01$, and Form B, F (3,366) = 5.47, $p < .01$. This effect was largely due to the considerable increase in judged ambiguity between the first and second blocks of 50 items in both inventories (Table 1). There was also a substantial decline in rated ambiguity for the last block of Form A, but not Form B.

TABLE 1.—Mean item ambiguity for each block of 50 items within the A and B forms of the 16PF.

Block	Form	
	A	B
1	.39	.35
2	.44	.42
3	.43	.40
4[a]	.37	.41

[a] includes only 37 items

The mean item ambiguity values for the scales of both 16PF forms are presented in Table 2. It can be seen that there was considerable agreement between the two forms, as reflected in the correlation between the rank-orders of the scales, $r_s = .67$, $p < .01$. Scale B, the intelligence scale was rated highly ambiguous on both forms; however, no scale was rated remarkably low on either the A or the B form. There were also three scales which tended to show some differences in ambiguity between the two forms; specificaly scales L (trust-suspiciousness), M (conventionality), and Q_1 (conservative-liberal). Both L and Q_1 were ranked higher on Form A than B, while the inverse was true for M; however, analysis of the variations indicated that the differences did not achieve statistical siignficance ($.05 < p < .10$ in each comparison).

TABLE 2.—Mean item abmiguity values for the factor scales from the A and B forms of the 16PF.

Scale	Form	
	A	B
A	.39	.38
B	.52	.55
C	.40	.34
E	.41	.40
F	.36	.34
G	.45	.48
H	.36	.37
I	.34	.37
L	.46	.38
M	.41	.49
N	.40	.45
O	.42	.38
Q_1	.53	.44
Q_2	.42	.39
Q_3	.37	.39
Q_4	.37	.32

On the A form, 80% of the items were rated more ambiguously by females than males, while on the B form the inverse was true as males had higher ambiguity ratings than females on 81% of the items. While a sign test applied to the proportions for each form indicated that the proportions differed significantly in each case ($p <$ .001), the differences between sexes in the magnitudes of the mean ambiguity ratings were small (two to five per cent) and not statistically significant. Moreover, the correlation between sexes on item ambiguity ratings was substantial for both forms, at .58 ($p < .01$) for the A form, and .65 ($p < .01$) for the B form.

The only factors in Form A with scores which showed significant positive correlations with the general tendency of a subject to rate an item ambiguous (as indicated by the proportion of items which a subject rated as having some degree of ambiguity) were *O* (guilt) and Q_4 (tension). Three factors *C* (ego-strength), *E* (dominance), and *I* (tough-tender) showed some tendency to be negatively related to ambiguity, however, the correlations did not achieve statistical significance (Table 3).

The relationships between subject ambiguity and scale scores for the B form were quite different than for the A form. Only factor *A* (aloof-outgoing) was found to correlate significantly with the tendency to rate Form B items as ambiguous, and this correlation was in the negative direction. Factor *H* (shy-bold) also showed a trend toward a negative correlation, however, the relationship was not sufficient to achieve statistical significance.

TABLE 3.—Intercorrelation between mean ambiguity values for *S*s and scores on 16PF factor scales.

Scale	Form	
	A	B
A	−.13	−.28*
B	.01	.15
C	−.24	.12
E	−.22	.01
F	−.17	.03
G	−.11	.05
H	−.10	−.21
I	−.21	.12
L	.15	.03
M	−.17	−.02
N	.15	−.04
O	.29*	.11
Q_1	−.10	−.18
Q_2	−.06	.10
Q_3	−.17	.01
Q_4	.29*	.02

*p <.05

The 16PF provides an opportunity to examine the relationship of items ambiguity to at least one aspect of scale validity, that of "direct concept validity" as discussed by Cattel, Eber, and Tatsuoka.[3] This validity measure describes the relationship of the inventory factor to the "pure" factor which it is supposed to measure. Using the values provided by these authors, the factor scales were rank-ordered in terms of increasing validity, and then correlated with the rank-order of the scales with respect to ambiguity. While the correlations of −.53 for the A form and −.36 for the B form suggested an inverse relationship, only the former value reached an accepted level of statistical significance ($p < .05$).

Another estimate of the relationship of ambiguity to the 16PF can be determined by comparing the rank-order of the equivalence coefficients for the A and B forms from each factor, to the order of scale ambiguity for each form. The resulting correlations were −.64 ($p < .01$) for Form A, and −.70 ($p < .01$) for Form B.

B. *MMPI*. The mean item ambiguity value for the MMPI was .30 (S.D. = .21); a value significantly lower than that of .37 (S.D. = .13)

obtained by Harris and Baxter,[5] t (208) = 2.81, $p < .01$. Similarly, the percentage of items judged ambiguous by a majority of the subjects was only 5.6 in this study, compared to the 15% reported by Harris and Baxter. The difference in proportions was also significant, X^2 (1) = 28.83, $p < .001$.

Even though there was a difference in level of ambiguity between this study and that of Harris and Baxter,[5] there was a high degree of correlation between the ambiguity values for individual items, $r = .74$ ($p < .001$).

As Harris and Baxter[5] had previously found, there were pronounced variations in level of item ambiguity as a function of item sequence within the MMPI (Table 4), F (3,1109) = 5.78, $p < .001$). As with the 16PF there was a trend toward an increase in rated ambiguity for the second block of 50 trials, however, this was followed by a considerable decrease across blocks 3 and 4. After a slight recovery, rating levels tended to be fairly constant until the last two or three blocks of items which returned to relatively low ambiguity ratings. That these sequence effects are generally characteristic of the MMPI is suggested by the correlation between ranked block ambiguity vaues for this and the Harris-Baxter administrations of .74 ($p < .01$).

TABLE 4.—Mean item ambiguity for each block of 50 items within the MMPI for both the present and Harris and Baxter (1965) administrations.

Block	Administration	
	Present	Harris and Baxter
1	.30	.42
2	.34	.41
3	.30	.36
4	.25	.34
5	.30	.36
6	.31	.38
7	.31	.39
8	.32	.39
9	.31	.36
10	.28	.33
11	.23	.31
12[a]	.25	.33

[a] includes only 16 items

The mean ambiguity value for the items comprising each MMPI scale are shown in Table 5. The scales with the least amount of ambiguity were L, F, and Hs, while scales Ma, K, *Welsh A*, Pd, and Si tended to have relatively high rated ambiguity levels. This ranking of scales was in good agreement with the ranking of the Harris-Baxter scale values, $r_s = .89$, $p < .01$.

TABLE 5.—Mean item ambiguity values for each MMPI scale for both the present and Harris and Baxter (1965) administrations.

Scale	Administration	
	Present	Harris and Baxter
L	.22	.33
F	.24	.31
K	.37	.45
Hs	.22	.31
D	.31	.41
Hy	.31	.41
Pd	.36	.43
Mf	.32	.35
Pa	.32	.41
Pt	.34	.43
Sc	.28	.39
Ma	.38	.46
Si	.35	.41
A	.37	.43
R	.30	.33
Es	.29	.36

As in the Harris and Baxter[5] study, men tended to rate MMPI items as being more ambiguous than did women. The difference between the percentage of items rated more ambiguous by men (66.9%) and the percentage rated more ambiguous by women (33.1%) was significant by a sign test ($p < .001$). However, as with the 16PF, the difference in magnitude of mean ambiguity values was only three percent, a nonsignificant difference. The correlation between males and females across all items was .75 ($p < .01$).

The correlations between subject ambiguity scores and the scores obtained on each MMPI scale are presented in Table 6. Three scales, Hs, Pt, and Sc, all of which were positively correlated to ambiguity, showed statistically significant correlations. In addition, the two scales D, and Es (a negative correlation) nearly achieved significance. In comparison, Harris and Baxter found significant positive correlations for scales Hs, Sc, Si and *Welsh A* and a significant negative correlation for the K scale. As in this study, the correlation which they obtained for the Es scale also approached significance.

TABLE 6.—Intercorrelation between mean ambiguity values for *S*s and scores on MMPI scales for both the present and Harris and Baxter (1965) administrations.

Scale	Administration	
	Present (N = 105)	Harris and Baxter (N = 74)[a]
L	.07	−.08
F	.04	.08
K	−.01	−.27*
Hs	.25*	.38**
D	.19	.16
Hy	.16	.15
Pd	.16	.03
Mf	.13	.00
Pa	.08	.20
Pt	.22*	.22
Sc	.21*	.25*
Ma	−.02	.10
Si	.05	.28*
A	.15	.25*
R	.08	−.07
Es	−.19	−.21

*$p < .05$
**$p < .01$
[a] Harris and Baxter included only the *S*s with 37 highest and 37 lowest ambiguity scores from their sample of 112 *S*s.

C. *16PF—MMPI Comparison.* The mean ambiguity value for the A form of the 16PF of .41 was significantly higher than the corresponding value of .30 obtained for the MMPI in this study, t (156) = 3.06, $p < .01$. Similarly, the ambiguity value of .40 for the B form was also found to differ significantly from the value for the MMPI, t (155) = 2.78, $p < .01$.

IV. Discussion.

The findings from this study suggest that many 16PF items are rated as having a substantial degree of ambiguity, and that judged ambiguity is at a considerably greater level that that associated with most MMPI items. This suggests some empirical support for Harsh's contention about the degree of ambiguity in 16PF items.

The 16PF scale which presented the most ambiguity, Factor *B* (intelligence), was the only factor for which more than one-half of the items exceeded an ambiguity rating of .50 for both Forms A and B. It was noted that ambiguity in this scale tended to increase as a function of item difficulty. It is therefore likely that ratings made on items in this scale were determined most directly by the degree of difficulty which *S* had in choosing the correct answer to relatively clearly worded questions, rather than primarily on the basis of uncertainty about the meaning of the questions. It is possible that this type of confounding in the ratings could have occurred with any item in the questionnaire; however, the differences in item structure between intelligence items (relatively clear wording and a single correct answer) and personality items (relatively imprecise wording with no specific correct answer) seems to account for this effect. In other words, it is probable that this departure from the intent of the instructions for making ratings of ambiguity was a significant factor only for intelligence items and not for personality items.

In some respects, the 16PF results raise questions about the comparability of the A and B forms. While there was a considerable degree of comparability between the mean ambiguity ratings for the majority of scales on the two forms, the discrepancies between the scales which most closely correlated with mean item ambiguity for *S*s suggest that these scales may not necessarily be measuring the same personality dimensions, or at least are measuring sufficiently different aspects of a given dimension to question the homogeniety of the underlying factor.

With respect to the MMPI, most of the findings from this study were consistent with those reported by Harris and Baxter.[5] While there was a significant difference between these two studies in overall level of ambiguity, the substantial item ambiguity correlations, the correspondence between sequence effects, and the general similarity of findings with respect to the relationships of scale scores to ambiguity suggest that the ambiguity characteristics described by these two studies are relatively stable inventory attributes. The difference in overall item ambiguity levels between the two studies seems most likely to have been a function of the single difference in rating procedure for the two studies. While Harris and Baxter[5] separated the item-answering and item-rating tasks by an interval of a week, subjects in this study answered and rated items concurrently. This suggests either that the concurrent procedure suppresses willingness to rate items as ambiguous, or that the degree of ambiguity perceived in an item is less when actually being answered than when it is

being considered "abstractly" for ambiguity. There is no evidence in these two studies to indicate which procedure yields the "truer" ambiguity values.

With respect to sex differences in ambiguity ratings, the results were marginal and mixed. On one hand ambiguity ratings by males were higher on more MMPI and 16PF-B items than ratings by females; a result consistent with the findings of Harris and Baxter.[5] On the other hand, female ratings were higher than male ratings on more items of the A form of the 16PF and, in addition, differences in mean ratings for the sexes on the various inventories were quite small and nonsignificant. Therefore, it would seem most appropriate to conclude that the sexes do not differ to any substantial degree in their tendency to perceive ambiguity in these kinds of situations. It might be speculated on the basis of the differences in proportions, that the sexes do tend to react differentially to ambiguity in the various inventories; however, the determinants of such differential reactions, if in fact they exist, are not readily apparent.

Examination of the intercorrelations between subject ambiguity values and scores on the various inventory scales suggest that indicators of positive emotional adjustment tend to be negatively correlated with ambiguity ratings, while scores associated with intrapersonal distress and defensiveness are positively related to these values. These findings are consistent with the results reported by Harris and Baxter[5] and support their conjecture that perception of ambiguity in situations may be related to anxiety levels and the ability of an individual to deal effectively with these feelings.

The finding that both validity and equivalency coefficients for the 16PF vary inversely as a function of ambiguity supports the contention that the utility of this type of assessment technique may be limited to some degree of item ambiguity. On the other hand, the ability of individuals to consciously distort their responses to present a biased personality picture may be inversely related to ambiguity. Therefore, it may well be that some item ambiguity is desirable, while too much or to little ambiguity may interfere with test utility. Determination of these effects will be especially important to the use of the techniques in screening applications, such as have been undertaken with air traffic controllers, since it is usually not possible to verify the accuracy of the assessments through follow-up interview procedures except for a very few individuals with significant scores.

REFERENCES

1. Anastasi, A.: *Psychological Testing*, New York, Macmillan, 1954.
2. Baxter, J. C., and V. L. Morris: Item ambiguity and item discrimination in the MMPI, JOURNAL OF CONSULTING AND CLINICAL PSYCHOLOGY, 32:-309–313, 1968.
3. Broen, W.E.: Ambiguity and discriminating power in personality inventories, JOURNAL OF CONSULTING PSYCHOLOGY, 24:174–179. 1960.
4. Cattell, R. B., H. W. Eber, and M. M. Tatsuoka: *Handbook for the Sixteen Personality Factor Questionnaire (16PF)*, Champaign, Illinois, Institute for Personality and Ability Testing, 1970.
5. Harris, J. G., Jr. and J. C. Baxter: Ambiguity in the MMPI, JOURNAL OF CONSULTING PSYCHOLOGY, 29:112–118, 1965.
6. Harsh, C. M.: 16PF Test: A review, In O. K. Buros (Ed.), *The Fourth Mental Measurements Yearbook*. Highland Park, N. J., Gryphon Press, 1953.
7. Karson, S. and J. W. O'Dell: Performance ratings and personality factors in radar controllers, FAA Office of Aviation Medicine Report No. AM–70–14, 1970.