RESULTS OF THE ESA STUDY ON PSYCHOLOGICAL SELECTION OF ASTRONAUT
CANDIDATES FOR COLUMBUS MISSIONS II: PERSONALITY ASSESSMENT

Klaus-Martin Goeters & Christoph Fassbender

German Aerospace Research Establishment (DLR)
Institute of Aerospace Medicine
Dept. of Aviation and Space Psychology
Sportallee 54, D-2000 Hamburg 63
Germany

ABSTRACT

A unique composition of personality assessment methods was applied to a
group of 97 ESA scientists and engineers.   This group is highly
comparable to real astronaut candidates with respect to age and
education.   The list of used tests includes personality questionnaires,
problem solving in groups as well as a projective technique.   The study
goals were: 1. Verification of psychometric qualities and applicability
of tests to the target group; 2. Search for culture-fair tests by which
multi-national European groups can be examined; 3. Identification of
test methods by which the adaptability of the candidates to the
psycho-social stress of long-duration space flights can be assessed.
Based on the empirical findings, a test battery was defined which can be
used in the selection of ESA space personnel.

INTRODUCTION

Space Station Freedom is planned as permanent research facility in the
earth orbit.   The European contribution to the system is the Columbus
Attached Laboratory.   It will allow astronauts to live and work under
the condition of microgravity, conducting experiments in material
sciences, fluid physics and life sciences.   Space Station Freedom
usually will be manned by international crews. Mission duration will be
extended up to 6 months.   Compared to the 1 week flights of the NASA
space shuttle, these are long-term missions which require astronauts
with an enhanced degree in motivation, interpersonal skills and stress
resistance in order to guarantee an optimal performance of the station
crew.   Loss of motivation, innergroup conflicts or the development of
anxieties would have a disastrous influence on reaching the mission
goals.   Therefore, the European Space Agency (ESA) funded a study on
"Definition of Psychological Testing of Astronaut Candidates for
Columbus Missions" (Goeters & Fassbender [5]), which was performed in
international cooperation. Partners of the German Aeroapace Research
Establishment (DLR) in Hamburg were the Universities of Reims and
Bergen. The importance of the assessment of personality factors is
emphasized by the fact that three of four work packages in the study
were dealing with such aspects.

Test methods employed in the selection of astronauts have to reliably
differentiate among persons in a preselected group as well.   Natural
scientists, engineers, and physicians are permitted to apply as ESA
astronauts for Columbus missions.   This group is characterized by a
specific intellectual level and also by a behavioral style which is
typical for scientists.   Most of the applicants do not suffer from any
mental or behavioral disorder, but are within the normal range of
personality. Therefore, the ESA test system has to make sensitive
differentiations in groups of people with a high cognitive level in the
normal range of personality traits in real situations of personnel
selection. When devising a new test system, all tests have to be
empirically verified with testees who are representative of later
applicants.   This verification is performed in order to receive
information about: test score distributions, reliability of scales,
factorial structure of test variables, and redundancy of test methods.
In this study an empirical check of culture fairness was also performed,
since ESA as an international organization accepts employees coming from
different European countries. An official ESA selection system must be

able to screen the astronaut candidates of different national heritage
in an unbiassed way. The inclusion of evaluation methods into the final
test battery to be administered in real astronaut selection was decided
on the basis of the found empirical information. The study was launched
in February 1990 and was completed in January 1991. Actual testing was
performed in Noordwijk, Netherlands. 97 ESA employees from European
Space Technology Centre (ESTEC) took part in the study as testees. The
empirical check was performed with ESTEC personnel because it is
supposedly representative for ESA astronaut candidates with respect to
age, to profession (scientists or engineers) and to national heritage
(from all ESA member states).

PERSONALITY ASSESSMENT METHODS

A diversity of questionnaires and objective tests such as problem
solving in groups or such as an active control of video presentation
(measurement of personal distance) as well as a projective test were
included in the study. If ever possible, the tests have been adapted
and translated into simple English and items have been controlled for
culture fairness. Such an adaption could not be made with methods taken
from the general test market, because one has to obey the copyright
which usually makes it impossible to change the test.

The various personality assessment methods which are covered by the
study are presented in Table 1.. The personality tests were supplemented
by a biographical questionnaire which collected descriptive data in
order to check the influence of such factors as age, gender, nationality
and the intention to apply in a real astronaut selection campaign. More
detailed biographical information about the candidates was not collected
because the anonymity of testees had to be protected. A test of English
(ENS) served as a control variable assessing the effects of language
knowledge on the test results.

Table 1. List of personality assessment methods and scales

Personality Questionnaires proposed by DLR (Hamburg):
    Temperament Structure Scales (TSS, see Maschke [10] and Goeters &
Timmermann   [6])   with   sub-scales   Achievement   Motivation   (ACH),
Aggressivity (AGG), Dominance (DOM), Empathy (EMP), Extraversion (EXT),
Mobility (MOB), Rigidity (RIG), Spoiltness (SPO), Emotional Instability
(STA), Vitality (VIT) and the control scale Openness (OPN).
    Fear Survey Schedule (FSS by Wolpe & Lang) with sub-scales
Agoraphobia (A), Fear of Blood, Injury, Illness and Death (B), Social
Fears (S), Fear of Sex and Aggression (SA) and Fear of Harmless Animals.
    State-Trait Anxiety Inventory (STAI, see Spielberger [11]) with the
sub-scales State Anxiety (SA) and Trait Anxiety (TA).

Personality Assessment Methods proposed by the University of Reims:
    Videotaped Interpersonal Distance Measure (VID, see Glimour &
Walkey [4]): Objective measurement of Tolerance of Approach (TOAP).
    Test of Decision Making (TD9, see Fourcade [3]) with 9
questionnaire scales Sincerety (SN), Impulsivity (IM), Taste for Games
of Chance (JE), Interpretative Projection (PR), Speed of Decision (DE),
Level of Aspiration (AS), Dynamism (DY), Tolerance of the Non-Control of
the Environment (CO) and Taste of Risk (RI).
    Matrix of Intra and Inter-Personal Processes in Groups (MIPG, see
Abraham [1]) with the variables Difference between Self and Ideal Image
(SI) and Difference between the Image of Self and Others (SO).
    Group Discussion (GD): Problem Solving in Groups (Assessment center
with rating of various behavior aspects incl. a Global Appreciation
(GLO))
    Defence Mechanism Inventory (DMI): Questionnaire with sub-scales
Turning Against Objects (TAO), Projection (PRO), Principalization (PRN),
Turning Against Self (TAS) and Reversal (REV).
Personality Assessment Methods proposed by the University of Bergen:
    Helmreich Battery: Personality Characteristics Inventory (HEB/PCI,
see Helmreich & Wilhelm [7]) with sub-scales Instrumentality (I),
Expressivity (E), Masculinity-Femininity (MF), Achievement Motivation
(DRIVEN), Impatience / Irritability (IMPAT5), Mastery (MAST),
Competition Orientation (COMP) and Work Orientation (WORK).
    Defence Mechanism Test (DMT, see Kragh [8]): Projective Test.

SUBJECTS

97 engineers and scientists of ESTEC participated in the study as
testees, thus forming a normation group for the applied psychological
assessment methods. The participation in the study was voluntary basis.
The age range of the subjects was from 24 to 38 (mean = 31). 82 testees
were male, 15 female. 30 testees were French, 23 testees Italian. All
other West European nationalities appeared less frequent. Therefore,
the nationalities had to be regrouped in order to reach adequate cell
frequencies. A meaningful regrouping was seen in the differentiation of
60 mediterranian (Italien, French, Spanish) versus 37 subjects of
northern national heritage (e. g. British, German, Scandinavian etc.).
The factor Nation (North vs. South) is refering to this differentiation.
27 testees declared "will also apply in a real astronaut selection
campaign"; 12 declared "might apply", but the majority of 58 declared
"will not apply".

RESULTS

The analyses of personality assessment methods, which are proposed by
DLR (see Table 1.), can be taken as representative for the basic
psychometrics which were of interest in the study. Thus this paper only
describes the empirical results of these methods. Thereafter in a meta-
analysis, the interrelation of all personality assessment methods as
listed in Table 1. is described and discussed. The meta-analysis gives
information concerning the content validity of the total set of
personality tests.

SCORE DISTRIBUTION

The mean, standard deviation, skewness and kurtosis of the various
subscales of the personality questionnaires TSS, FSS and STAI were
computed. Critical values are standard deviations, which are small
(less than 10% of the mean), and coefficients of skewness and kurtosis,
which exceed   ±1. In the TSS no such critical values are obtained
indicating that the distribution of scores on the subscales of this test
are not very different from a normal distribution without critical
distortions. This is a good basis for the differentiation between
candidates over the whole range of measurement. In the Fear Survey
Schedule FSS, two scales show critical values of skewness and kurtosis
while the rest of the scales seem to be satisfactory. The critical
scales are the subscales Agoraphobia and Fear of Harmless Animals. In
these scales only a few candidates admit anxieties of this sort. On the
other three scales of the FSS, namely Fear of Blood, Injury, Illness
and Death, Social Fears, and Fear of Sex and Aggression the score
distributions were not critical. In the STAI the State Anxiety scores
show a distorted distribution, while the scores on the Trait Anxiety
scale are well distributed. Distortions in the score distribution on
the State Anxiety scale can be neglected for selection purposes,
because in selection one is usually more interested in the trait
structure than in the actual state, which does not provide as good a
predictability as the trait component. The distortions in the FSS
subscales Agoraphobia and Fear of Harmless Animals require that these
scales have to be reworked in order to achieve a better sensitivity of
the scores which is necessary for its use in the astronaut candidate
population. Especially items indicating subliminal tendencies of
agoraphobia and fears of harmless animals should be included in these
scales in order to receive a continuous differentiation.

RELIABILITY ESTIMATES

Generally, more ambivalance exists in personality data than in
performance scores. Therefore it is hard to obtain a level of
reliability in personality assessment that is similar to the one for
aptitude tests. But an absolute minimum level of reliability exists for
personality assessment techniques as well. Reliability coefficients
below 0.60 are only expected for undeveloped scales (Lienert [9], p.
246). A reliability of 0.60 can be seen as the absolute minimum for
questionnaire scales when used for personnel selection. Reliabilities
between 0.60 and 0.70 only allow the interpretation of extreme scores
(e.g. 1 and 2 or 8 and 9 in the STANINE system or equivalent scores in
other standard distributions), because all other STANINE scores (3 to

7)   fall into the confidence interval of the mean.   Only if
reliabilities exceed 0.70, a more precise differentation can be made.
Therefore,   acceptable questionnaire reliabilities fall into the
interval 0.70 to 0.80.   Nevertheless, reliabilities above 0.80 are
desirable, but are often hard to receive.

The reliability estimates for the questionnaire scales of TSS, FSS and
STAI were determined by Cronbach's alpha providing an indication of the
internal consistency of the scales.   No reliability falls below the
absolute minimum value of 0.60. The reliability coefficients range from
.69 (FSS_A) to .88 (TSS_DOM) with a median of .80. The scales, which
turned out to be critical with respect to their score distributions,
surpassed the critical score in reliability.   That means that at least
extreme scores in these scales are interpretable with some confidence.
With the exception of Agoraphobia, all scales show either acceptable
reliabilities between 0.70 and 0.80 or even good reliabilities between
0.80 and 0.90. No particular scale has to be excluded from the test set
for lack of reliability.

DEPENDENCY ON SUBJECT CHARACTERISTICS

Concerning the dependency of test data on subjects' characteristics the
following   factors   are   checked: language   bias   of   English,
culture—fairness (national background), gender, age and the intention
to apply for an astronaut position. Especially the first two components
are essential factors in this study.   The test set should be free of
language and nationality effects, since it is designed for application
in all ESA member states.   Personality questionnaires which are
completely language—dependent may result in strong language effects.
This is also true for nationality effects, because many items in
personality questionnaires are referring to habits,   which can be
influenced by national customs.   The check of gender and age effects is
normal routine in the analysis of psychological tests.   A check of the
intention for application is included in the study,   because the
hypothesis was that motivation effects could be an important factor
influencing the results.

Only few significant effects exist concerning the influence of age,
gender, and intention for application. The TSS control scale Openness
(TSS_OPN) shows significant gender and age effects.   Males and older
subjects react more openly in the questionnaire than females and younger
testees.   Since these effects appear in the control scale,   no
differential norms or other countermeasures are required in this case.
The TSS scale Mobility (TSS_MOB) also shows a significant effect: the
older the subjects the less they score on the mobility scale. From the
scale content one can conclude that older subjects tend to describe
themselves as less mobile and less risk—taking than younger ones.   Such
a dependency seems to be a natural development which has not to be
corrected by differential age norms. Another significant effect exists
for the FSS scale HA: Females report significantly more anxieties with
respect to harmless animals than males.   This corresponds to a common
stereotype of female behavior and thus may represent "real" behavioral
differences which must not be corrected by differential gender norms.
No other significant effects appear. Therefore, one can conclude that
the personality scales which are under discussion here are more or less
robust   against   influences   of   age,   gender,   and   motivation   for
application.   No specific correction for these factors seem to be
necessary for the administration of these scales in the real selection
situation.

Since knowledge of English and national heritage are crucial factors to
be analysed, and since both factors are correlated to some extent, an
analysis of covariance is performed in order to split up the effects of
knowledge of English and national heritage.   In this analysis of
covariance the knowledge of English was used as the covariate.   The
results of this analysis of covariance are shown in Table 2.   What can
be seen from this table is that knowledge of English is the most
dominant   effect   while   nationality   alone   seems   to   be   neglectible.
Knowledge of English is especially dominant for the FSS and STAI, while
all subscales of the TSS are robust against English.   More than 50 per
cent of all scales of FSS and STAI are significantly dependent on
knowledge of English. This fact is easy to explain, as the FSS and STAI

were developed by native speakers of English. In contrast, the TSS
scales were constructed by psychologists who are not native speakers of
English. They automatically avoid strange vocabulary which is difficult
to understand for non-native speakers. Such subjects, whose native
language is not English, constitute the majority of the ESA testees.
As the TSS scales are independent with respect to different nationality
and language backgrounds, it can be recommended that the TSS can be
directly included in the final ESA test battery. The FSS and the STAI
cannot be directly used in an international group of testees. Half of
the FSS and STAI scales should be adapted to these requirements by
revising the test items. New item substitutes have to be formulated
which avoid rare and difficult English expressions.

Table 2.   Personality questionnaire data: Analysis of variance for
                different groups of NATION (North vs. South) with ENS
                (English Test Score) as covariate

	F – Values			F – Values	
Scale	ENS	NATION	Scale	ENS	NATION
TSS_ACH	.022	.109	FSS_A	4.106*	.806
TSS_AGG	.972	2.212	FSS_B	1.519	2.339
TSS_DOM	.383	.088	FSS_S	2.311	.014
TSS_EMP	.087	.009	FSS_SA	4.011*	.233
TSS_EXT	.276	1.387	FSS_HA	3.206	2.421
TSS_MOB	.007	1.246			
TSS_RIG	3.676	.780	STAI_SA	6.875**	.335
TSS_SPO	.381	.407	STAI_TA	14.595**	2.624
TSS_STA	1.733	.761			
TSS_VIT	.109	2.802	Significant effects:		
TSS_OPN	.216	1.760	* < .05; ** < .01		

META–STRUCTURE OF PERSONALITY ASSESSMENT

All personality scales as listed in Table 1. were intercorrelated in
order to identify redundancies and the main components of this set of
data.   The matrix of correlations between all personality measurements
is shown in the appendix of the original study report. The following
statements can be derived from these correlations: Many of the
significant correlations do not reach the level of parallel test
reliability. It can be concluded that there is enough specific variance
in many cases, which represents different facettes of personality. The
main exception from the fact is the high level of intercorrelations of
the group discussion variables. All of the ratings received from this
assessment situation are so highly correlated that they refer to the
same content, although different labels were used. It seems that the
different group discussion ratings   can be reduced to one component
(e.g. variable GD_GLO),   which can be described as positive versus
negative behavior in the group situation. When trying to identify
significant correlations between group discussion ratings and other
variables,   positive correlations with the subscales Extraversion and
Spoiltness of the TSS and negative correlations with the subscale
Impatience of the Helmreich battery are found.   This is plausible:
People, who are extraverted   (Spoiltness is also a subdimension of
extraversion)   tend to be more lively in social interaction than
introverted subjects, thus stimulating the group activity. On the other
hand people who tend to be impatient, often appear disturbing in social
situations. The significant correlations of some TSS and HEB subscales
with "real" behavior in the group discussion can be seen as an
indication of the validity of these questionnaire scales.   It is
prognostic validity because the questionnaire data is always collected
before the group discussion (at least 1 day beforehand). At the
beginning it was speculated that the DMT would not be related to any
other personality measure. This assumption is based on earlier studies
employing the DMT. Unexpectedly, the Defence Mechanism Test (DMT) is
correlated with a number of personality scales such as with subscales of
the FSS, the TSS, Group Discussion, HEB, DMI, TD9, and MIPG.   A
description and discussion of these correlations led to the
interpretation of a life style which is dominated by sportive and risk

seeking behavior.

Factor analyses were computed for the identification of the meta-structure of the total set of personality assessments. It was not possible to perform a factor analysis for the complete set of variables, because this surpassed the capacity of the computer used. Therefore, different strategies were applied in order to reduce the number of variables. From the Group Discussion only the Global Appreciation (GD_GLO) was taken because of the high redundancy in this method (see above). Seven of the nine TD9 scales were excluded since a preliminary factor analysis revealed a test specific factor indicating also some redundancy among various TD9 scales. Only those two TD9 scales remained in the final factor analysis which showed the most substantial loadings on the test specific factor, namely TD9_PR and TD9_RI. From the FSS only the total score and from the STAI the trait scale were used. All other questionnaire scales (see Table 1) as well as the scores of the MIPG (self image) and the variable Tolerance of Approach (VID_TOAP) of the Videotest could be included into the final factor analysis. Various summary scores of the DMT were first included in this analysis as well, but they finally had to be excluded because the iteration program of the factor analysis did not reach a difinite solution.

The reduced data set of all personality assessment methods excluding the DMT was factoranalyzed according to the method of principal components. The final factor solution resulted in seven varimax-rotated factors. Based on loadings greater than .40, the factor structure leads to the following interpretations:

Factor 1 defined by the variables (substantial loadings in brackets): DMI_TAO (.85), DMI_REV (-.76), DMI_PRO (.71), DMI_PRN (-.66), TSS_AGG (.59), TSS_OPN (.56).
Scales of aggression and conflict treatment load on this factor. So, this factor can be labled Aggressive Behavior.

Factor 2 defined by the variables (substantial loadings in brackets): STAI_TA (.73), FSS (.73), TSS_STA (.65), VID-TOAP (.62), MF (-.58), TD9_PR (.52), IMPAT5 (.49), DMI_TAS (.49).
Scales of anxiety, emotional instability, and irritability load on this factor. Therefore, this factor is clearly to be identified as the Neuroticism factor which is described by Eysenck [2]. It is of interest that the Videotest score "Tolerance of Approach" is highly correlated with this factor, too. High scores in the Videotest indicate that the subject feels very uncomfortable and irritated while approached. It seems as if the the approach of others creates anxieties.

Factor 3 defined by the variables (substantial loadings in brackets): DRIVEN (.82), TSS_ACH (.74), MAST (.60), I (.57), TSS_RIG (.53), COMP (.48).
Scales of Achievement Motivation and Endurance load on this factor and the factor can be labled respectively.

Factor 4 defined by the variables (substantial loadings in brackets): TD9_RI (.75), TSS_MOB (.75), TSS_VIT (.62).
This factor includes scales of Risk Taking and Sensation Seeking. Factor 4 appears as a general behavior component.

Factor 5 defined by the variables (substantial loadings in brackets): TSS_EMP (.68), E (.67).
Scales of Empathy load on Factor 5 and it can be labled respectively.

Factor 6 defined by the variables (substantial loadings in brackets): TSS_SPO (.68), GD_GLO (.51), WORK (.49), TSS_EXT (.48).
Factor 6 represents the aspect of extraversion with several subscales. It can be interpreted as the Level of Social Interaction. It is of fundamental interest that this factor includes questionnaire as well as real life data (assessment of group discussion behavior: Variable GD_GLO).

Factor 7 defined by the variables (substantial loadings in brackets): MIP_SI (.85), MIP_SO (.67), TSS_DOM (-.42).
It seems that Factor 7 represents a Negative Self-Concept which is either indicated by the difference between the self and the ideal

concept or between the self concept and the concept, how others see the subject. Factor 7 as a negative self-concept correlates positively with the DMI scale "Turning against Self" (prior factor analysis) or negatively with the TSS scale "Dominance" (present analysis). Both correlations are meaningful with respect to the interpretation of the factor.

The extracted factors are clearly identifiable general personality components. The factor structure includes aspects of work orientation (Factor 3 "Achievement Motivation and Endurance"), of stress resistance (Factor 2 "Neuroticism", Factor 4 "Risk Taking and Sensation Seeking"), and of social interaction (Factor 6 "Level of Social Interaction", Factor 1 "Aggression", Factor 5 "Empathy"). Although Factor 7 is mainly concerned with the individual self concept, this factor may also be grouped under social interaction, since the characteristics of the self-concept can directly influence the behavior in groups.

The factor solution is exhaustive in the sense that all the included variables show substantial loadings. The solution is unequivocal because the variables mostly load on one factor, only in few cases on more than one. Scales of the TSS load on each factor. This indicates that the TSS is a very broad method. It completely covers the meta-structure of the intended diagnostic information. The second complete method is the HEB. Scales of the HEB show substantial loadings on 5 factors. Loadings on Factor 4 "Risk Taking and Sensation Seeking" and on Factor 7 "Negative Self-Concept" are missing. Some HEB scales are not unequivocal since they show substantial loadings on more than one factor. All other diagnoctic methods are specifically focussed on only a limited number of factors.

FINAL TEST BATTERY

Based on the presented empirical results a variety of personality assessment methods were chosen as constituents of the final test battery by which future ESA astronaut applicants will be screened: TSS and HEB as questionnaires, GD as an assessment center technique and the DMT as a projective test. Despite its unfavourable psychometrics the last test was chosen because this method may detect a critical disposition with respect to behavior under extreme stress.

REFERENCES

[1]   Abraham, A. (1972). Manuel d˜instruction du MIPG. Paris: Editions Scientifiques et Psychotechniques.
[2]   Eysenck, H.J. (1974). Dimensions of personality. London: Routledge & Kegan Paul.
[3]   Fourcade, J. (1975). Une approche quantitative de 1˜influence de la personnalite sur 1˜activite decisionelle: Un inventaire de decisions individuelles. Travaux Scientifiques CRSSA.
[4]   Glimour, D.R. & F.H. Walkey (1981). Identifying violent offenders using a video measure of interpersonal distance. Journal of Consulting and Clinical Psychology, 49, 287-291.
[5]   Goeters, K.M. & Ch. Fassbender (1991). Definition of psychological testing of astronaut candidates for Columbus missions. European Space Agency ESA 8730/90/NL/IW.
[6]   Goeters, K.M. & B. Timmermann (1991). The construction of personality questionnares for selection of aviation personnel. DLR- FB 91-18.
[7]   Helmreich, R.L. & J.A. Wilhelm (1989). Validating personality constructs for pilot selection. Austin / Ames: NASA/UT Technical Memorandum 89-3.
[8]   Kragh, U. (1960). The Defence Mechanism Test: A new method for diagnosis and personnel selection. Journal of Applied Psychology, 44, 303-309.
[9]   Lienert, G.A. (1969). Testaufbau und Testanalyse. Weinheim: Beltz.
[10]   Maschke, P. (1986). Temperament-Struktur-Skalen (TSS): Testmanual. DLR-FB 86-58.
[11]   Spielberger, C.D. (1983). State-Trait Anxiety Inventory. Palo Alto: Consulting Psychologists Press.