Factor Analysis on the Characteristics of Occupation

Bhina Patria
inparametric@yahoo.com
http://www.inparametric.com

Introduction

Factor analysis was first used in 1904 by Charles Spearman, a Psychologist from United Kingdom. Spearman used factor analysis in his models of human intelligence. With factor analysis he developed theory that variety of cognitive test could all be explained by one single factor which called *g* factor. Later on Raymond Cattell, psychologist from US, used factor analysis in most of his research on intelligence which lead to the development of his theory on Fluid and Crystallized Intelligence. 16 Personality Factor (16PF) test is also one of Raymond Cattell's research results based on factor analysis.

Factor analysis is quiet different to other statistical analysis method. Many statistical analysis methods are used to investigate the relation between independent and dependent variables. In contrast, factor analysis is used to study the patterns of relationship among many dependent variables, with the goal of discovering something about the nature of independent variables that affect them. The tricky part is that those independent variables were not measure directly. Thus answers obtained by factor analysis are necessarily more hypothetical and tentative than is true when independent variables are observed directly (Darlington, 1997).

Factor analysis in psychology is most often associated with intelligence research. Nevertheless, the use of factor analysis can be observed in other domain of psychology such as personality, attitudes, etc. The trait theorists in psychology have been used factor analysis to measure personality traits. The Extraversion-Introversion and the Neuroticism traits by Eysenck and 16 Personality Factor questionnaires (16 PF) by Cattell are some of the examples of implementation of Factor Analysis (Field, 2000). Furthermore, factor analysis is used widely in other social sciences, education, business fields, biological science, etc.

The main applications of factor analytic techniques are: (1) to *reduce* the number of variables and (2) to *detect structure* in the relationships between variables, that is to *classify variables*. Therefore, factor analysis is applied as a data reduction or structure detection method (Statsoft, 2003). Reducing the number of variables means reducing complexity of data, thus it will be easier to analyze the data.

CHEERS data on Characteristics of Occupation

The data used in this paper is based on the data collected in CHEERS. CHEERS (Careers after Higher Education: a European Research Study) was a research project conducted from 1998 to 2000. This survey was covering 3000 graduates each from 9 countries in the European Region, Norway, Czech Republic and Japan. The study is focused on the relationship between higher education and employment four years after graduation. Several themes were carried on the survey such as: socio-biographic background, study paths, transition from higher education to employment, early career, links between study and employment, job satisfaction and perspectives of the graduates on higher education (INCHER, 2001).

The data analysed in this paper is taken from section G of the questionnaire which is the section of Work Orientations and Job Satisfaction. The answer of the question G3A will be used in the analysis. Question G3: How important are the following characteristics of an occupation for you personally (A) and to what extent do they apply to your current professional situation (B)? *If you are not employed please answer only (A).* The scale of answer is from 1 = 'Very important' to 5 = 'Not at all important'. The complete characteristics of occupation can be observed in Table 1. From the total CHEERS data, 2000 cases were selected by Select Cases function in SPSS.

Table 1 Descriptive statistics on the importance of occupation characteristics

	N	*Mean*	*SD*
Largely independent disposition of work (G3AINDEP)	1966	1.92	.85
Opportunity of undertaking scientific/scholarly work (G3ASCIEN)	1962	2.89	1.26
Clear and well-ordered tasks (G3ACLEAR)	1965	2.45	1.05
Possibilities of using acquired knowledge and skills (G3ASKILL)	1965	1.79	.82
Job security (G3ASECUR)	1965	1.99	1.00
Social recognition and status (G3ASTATU)	1962	2.73	1.07
Opportunity of pursuing own ideas (G3AOWNID)	1962	1.84	.81
Good social climate (G3ACLIMA)	1965	1.46	.66
Opportunity of pursuing continuous learning (G3ALEARN)	1962	1.74	.84
High income (G3AINCOM)	1964	2.35	.93
Chances of (political) influence (G3AINFLU)	1961	3.03	1.21
Challenging tasks (G3ACHALL)	1962	1.97	.87
Good career prospects (G3ACAREE)	1963	2.25	.96
Enough time for leisure activities (G3ALEISU)	1965	1.93	.91
Co-ordinating and management tasks (G3AMANAG)	1963	2.69	1.03
Possibility of working in a team (G3AWTEAM)	1965	2.41	.98
Chance of doing something useful for society (G3AUSEFU)	1959	2.35	1.07
Variety (G3AVARIE)	1963	1.86	.82
Good chances of combining employment with family tasks (G3AFAMIL)	1959	2.14	1.16
Valid N (listwise)	1911		

Note: Question G3. How important are the following characteristics of an occupation for you personally (A) and to what extent do they apply to your current professional situation (B)? *If you are not employed please answer only (A).* The scale of answer is from 1 'Very important' to 5 'Not at all important'.
Source: CHEERS.

Factor Analysis

This paper utilized SPSS 13 for conducting factor analysis on the CHEERS data. The factor analysis menu can be accessed by mouse click from Analyse > Data Reduction > Factor (see Figure 1 in Appendix), however in this paper analysis will conducted by SPSS Syntax command. The complete SPSS Syntax command can be observed in the Appendix.

After running the analysis, the first table produced is Correlation Matrix. First observe the Determinant of matrix which is stated in the bottom of the table. The Determinant of matrix is important for testing for multicollinearity or singularity. The determinant of the correlation matrix (R-matrix) should be greater than 0.00001 (Field, 2000). In this analysis the value of R-matrix determinant is 0.026 (see Appendix), therefore it is proven that multicollinearity is not a problem for these data. The next step is checking the correlation pattern of the relationship between all of the variables. This step is important because in factor analysis variables which have too small correlation (or not significantly correlate) should be eliminated first. The same procedure should be applied to variables which correlate perfectly. Variables which correlate perfectly will also cause a singularity or multicollinearity problem. Therefore, one need to scan all significance value and identify variable which have the majority of values greater than 0.05. Afterwards one need to scan the correlation coefficients and look for values greater than 0.9.

From the CHEERS R-matrix we can observe that variable GA3CLEAR (Clear and well ordered tasks) does not significantly correlate ($p > 0.05$) with other variables (G3AINFLU, G3ACHALL, G3AMANAG, and G3AVARIE). As suggested by Field (2000), this variable should be omitted first before running the analysis again.

The next important output table is the KMO statistics. The KMO statistics represents the ratio of the squared correlation between variables to the squared partial correlation between variables. The value varies between 0 and 1. A value of 0 indicates that the sum of partial correlation is large relative to the sum of correlations, indicating diffusion in the pattern of correlation (hence, factor analysis is likely inappropriate). A value close to 1 indicates that patterns of correlations are relatively compact and so factor analysis should yield distinct and reliable factors (Field, 2000). For the CHEERS data the value is 0.79 (Table 2) which according to Kaiser (in Field, 2000) is in the good category. So, it is appropriate to use factor analysis in the CHEERS data.

Table 2
KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy	.795
Bartlett's Test of Sphericity	
Approx. Chi-Square	6685.591
df	153
Sig.	.000

The KMO statistic can be calculated for multiple variables and also for individual variable (Kaiser (1970) in Field, 2000). The KMO values for individual variables are produced on the diagonal of the Anti-Image Correlation Matrix. The values of the correlation should be above 0.5 for all variables. If there are values below 0.5 then removing the variable is recommended. For the CHEERS data all of the value is above 0.5 (see Appendix).

Table 3
Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.99	20.98	20.98	3.99	20.98	20.98	2.65	13.92	13.92
2	1.82	9.56	30.54	1.82	9.56	30.54	2.13	11.21	25.13
3	1.64	8.64	39.17	1.64	8.64	39.17	1.91	10.04	35.17
4	1.38	7.24	46.41	1.38	7.24	46.41	1.80	9.45	44.63
5	1.26	6.62	53.03	1.26	6.62	53.03	1.60	8.41	53.03
6	0.95	5.00	58.03						
7	0.84	4.44	62.47						
8	0.77	4.05	66.52						
9	0.74	3.87	70.39						
10	0.71	3.72	74.10						
11	0.68	3.60	77.70						
12	0.63	3.33	81.03						
13	0.61	3.18	84.21						
14	0.57	3.01	87.22						
15	0.54	2.82	90.04						
16	0.53	2.77	92.80						
17	0.51	2.66	95.46						
18	0.45	2.38	97.84						
19	0.41	2.16	100.00						

Extraction Method: Principal Component Analysis.

The Total Variance Explained (Table 3) explained the eigenvalues associated with each factor (linear components) before extraction, after extraction and after rotation (Field, 2000). Before extraction there are 19 components which are all variables listed. From the initial eigenvalues from the CHEERS data factor 1 explains 20.98 % of total variance. It is normal to have a large amounts of variance by the first factor then much smaller in the next factor (factor 2 explain only 9.56% of the variance). Under the Extraction Sum of Square Loadings only factor with eigenvalues bigger than 1 is listed, the result is only 5 factors. In the last part of the table the eigenvalues of the factor after rotation are displayed. Rotation of the factor axis has an effect which is optimizing the factor structure. The consequence for this data is that the relative importance of the four factors is equalized (Field, 2000). Before rotation, factor 1 accounted 20.98 % of all variance while other factor were much smaller (9.56%; 8.64%; 7.24%; and 6.62%). After rotation, the factor 1 account for only 13.92 % of variance, not a big difference from the others (11.21%; 10.04%; 9.45%; and 8.41%).

The last output from factor analysis on the CHEERS data is the Rotated Component Matrix (Table 5) which is a matrix of the factor loading for each variable on to each factor. Some of the matrix cells in this table are empty because previously in the analysis it was set up not to display values less than 0.4 (see

Figure 1). Steven (in Field, 2000) was the one who suggested setting a cut point of 0.4 factors loading for each factor. Factor loading higher than 0.4 considered appropriate for interpretative purposes.

Table 4
Rotated Component Matrix[a]

	Component				
	1	2	3	4	5
Challenging tasks (G3ACHALL)	.743				
Largely independent disposition of work (G3AINDEP)	.636				
Opportunity of pursuing own ideas (G3AOWNID)	.629				
Opportunity of pursuing continuous learning (G3ALEARN)	.564				
Variety (G3AVARIE)	.553				
High income (G3AINCOM)		.773			
Social recognition and status (G3ASTATU)		.701			
Good career prospects (G3ACAREE)		.643			
Enough time for leisure activities (G3ALEISU)			.728		
Good social climate (G3ACLIMA)			.632		
Job security (G3ASECUR)		.431	.545		
Good chances of combining employment with family tasks (G3AFAMIL)			.522		
Chance of doing something useful for society (G3AUSEFU)				.622	
Possibility of working in a team (G3AWTEAM)				.604	
Chances of (political) influence (G3AINFLU)				.571	
Co-ordinating and management tasks (G3AMANAG)		.411		.546	
Opportunity of undertaking scientific/scholarly work (G3ASCIEN)					.775
Possibilities of using acquired knowledge and skills (G3ASKILL)					.611

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
[a]Rotation converged in 11 iterations.

From this Rotated Component Matrix, it is identified that there are 5 factors and most of the variables correlate highly to only 1 factor. Only two variables are highly correlating with two variables which is variable Job security (G3ASECUR) and Co-ordinating and management tasks (G3AMANAG).

The next step of analysis is to investigate the content of each variable to find the common themes that represent each factor. The Job Characteristics Model (JCM) by Hackman and Lawler (Gibson, Invancevich, & Donnelly, 2000) will be used as a basic reference for the common themes of the factor. Variables which load highly to factor 1 appear to relate with the content of the occupation and based on the JCM this factor can be labelled as *Autonomy* factor. The variables which correlate highly on second factor are related to financial and status named as *Income and status* factor. The third factor is related to the opportunities to engage in social activities, this factor can be named as *Social opportunities* factor. The fourth factor is related to the extent that the occupation affects the well being of others; therefore this factor is labelled as *Task significance* factor. And the variables that load higher to the fifth factor are related to the possibility of using knowledge and undertaking scientific work, this factor is labelled as *Learning opportunity* factor.

After running factorial analysis there will be new variables in the data set which have names begin with FAC and followed by number of factor that was produced. With the CHEERS data the new variables are FAC1_6; FAC2_6; FAC3_6; FAC4_6; and FAC5_6. In order to make it easier in further analysis, renaming and labelling the variable according to the content was conducted. The new variable name is AUTO (Autonomy); INSTAT (Income and status); SOCOPT (Social opportunities); TASK (Task significance); and LEARN (Learning opportunity) with variable labels according to the factor name.

Discussion

In factor analysis the last step, finding the common themes that represent the all variable in the factor, is the most difficult part. In this step some subjectivity will take part. Hence, the same output analyse by two different researcher will yield a different labelling on the factors. Using one underlying theory to determine the factor is very important in this step in order to reduce this subjectivity. In this analysis the JCM (The Job Characteristics Model) is used as basic underlying theory to interpret the factor.

The other problem occurred when one (or more) variable has high correlation values to more than 1 factor. Some researchers directly delete this variable and then running the analysis again. This step in taken based on the belief that each factor is independent and therefore they are not supposed to be related to each other at any means. Each factor should not share the same variable. In this paper another approach is used to analyse this matter. Nevertheless, in this paper the variable which has high correlation to two factors (G3ASECUR and G3AMANAG) is not deleted. Variable which has a high correlation to two factor means that they have big account in forming those factors, thus deleting this variable means reducing the reliability of the measurement in the theoretic level. Furthermore, the fact that all factors are not correlating to each other, based in the correlation analysis (see Correlation Table in Appendix), is a proof that each factor does not have any relation to other factor even when they share the same variable.

Based on the statistical computation methods, variable G3ASECUR (Job security) should be classified in factor 3 instead of factor 2 because it has higher correlation to factor 3. However considering the content of the variable and the similar underlying theme, this variable is considered as the aspect of factor 2

Income and status. While the classification of variable G3AMANAG (Co-ordinating and management tasks) to factor 4 rather than factor 2 is merely based on the correlation value (0.546 to factor 4 versus 0.411 to factor 2). The complete factor of Occupation characteristics and its aspects can be observed in table 5. The factor data on the first 25 subject can be observed in table 6 (see Appendix).

Table 5. Characteristics of Occupation

Factor 1: Autonomy
- Challenging tasks
- Largely independent disposition of work
- Opportunity of pursuing own ideas
- Opportunity of pursuing continuous learning
- Variety
Factor 2: Income and status
- High income
- Social recognition and status
- Good career prospects
- Job security
Factor 3: Social opportunities
- Enough time for leisure activities
- Good social climate
- Good chances of combining employment with family tasks
Factor 4 Task Significance factor
- Chance of doing something useful for society
- Possibility of working in a team
- Chances of (political) influence
- Co-ordinating and management tasks
Factor 5 learning opportunities
- Opportunity of undertaking scientific/scholarly work
- Possibilities of using acquired knowledge and skills

Reference

Darlington, R., B. (1997). *Factor Analysis*. Retrieved 8th February 2007 from http://www.psych.cornell.edu/Darlington/factor.htm

Field, A. (2000). *Discovering Statistics Using SPSS for Windows*. London: SAGE Publication.

Gibson, J. L., Invancevich, J. M., & Donnelly, J. H. (2000). *Organizations: Behavior, Structure, Processes*. US: Irwin McGraw-Hill.

INCHER. (2001). *CHEERS Higher Education and Graduate Employment in Europe - European Graduate Survey*. Retrieved 8th February 2007 from http://www.uni-kassel.de/wz1/tseregs.htm.

StatSoft. (2003). *Principal Components and Factor Analysis*. Retrieved 8th February 2007 from http://www.statsoft.com/textbook/stfacan.html.

Appendix

Figure 1 Factor Analysis with SPSS

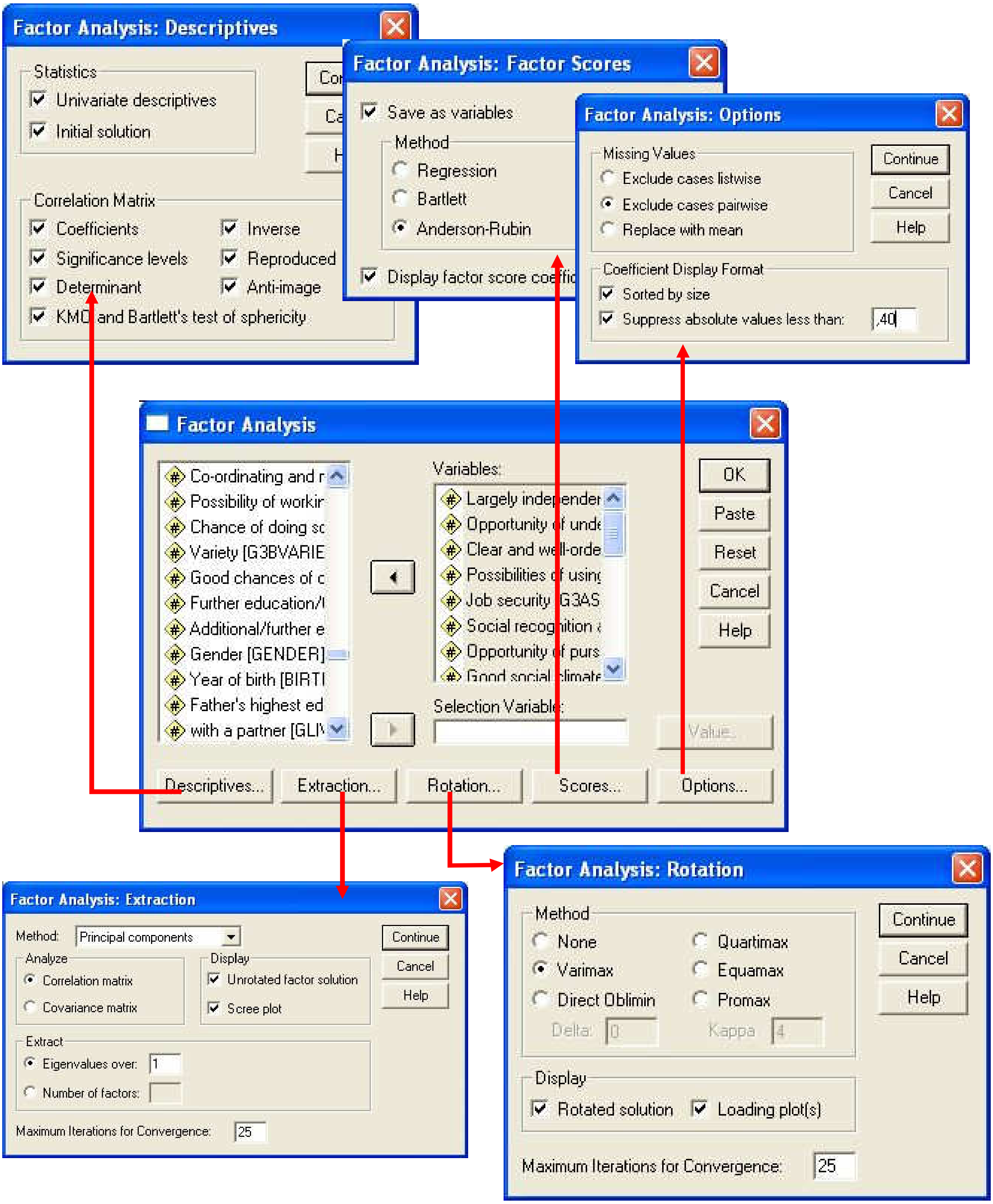

Table 6 Case Summaries[a]

	AUTO	INSTAT	SOCOPT	TASK	LEARN
1	-1,01793	-1,40649	-,27086	-,58486	-1,21488
2	-,71021	-,32751	-,30500	-,91313	-1,58312
3	-1,31709	-,67050	,08566	,35636	1,40392
4	,29247	-,34042	,84519	,32152	-1,87967
5	1,66527	-,18675	-,08998	-,46560	-,19384
6	,56767	-,57020	,68423	-,21218	-,40504
7	-1,06497	-1,73525	,26961	,25226	-,45716
8	,15822	1,81677	-1,05058	-,06112	-1,90730
9	-,22799	1,37618	-,59638	-,47199	-,31202
10	-,31429	-,18848	,66596	-1,79502	-,14801
11	-,53312	-1,12656	,76055	-,35702	1,04078
12	-1,75022	-,00136	,38715	-,33006	,48762
13	-,33855	1,35394	,00196	-1,64058	1,08278
14	-,41466	-,66427	-,29921	-1,04687	-,22102
15	-1,07907	-1,51108	,05948	-1,29068	-,70672
16	-1,25551	1,45189	,40196	-1,07325	-,94851
17	-,20977	-2,01525	-,03093	,58632	-,20355
18	,40919	-,67636	-,28782	-,92884	1,03653
19	,15808	-,11274	-,95528	-,89223	-,84558
20	,81761	-,03041	1,34848	-,24867	1,01613
21	-,03950	3,70655	-1,97291	,73233	-1,21879
22	-,25874	-,93662	-,01602	-1,05637	-,65423
23	-1,01564	-,38649	,66067	,56545	-1,21915
24	-1,73404	-,76484	1,57870	-,61013	-,47285
25	-,35301	-,55077	,47205	-1,51512	,82501
Total N	25	25	25	25	25

Note. [a]Limited to first 25 cases.

SPSS Syntax Command

```
*OPEN DATA.
Get file = 'd:\my
document\mahe_modules\13modul7_2_finished\grad_mahe09.sav'.

*DESCRIPTIVES STATISTICS.
Formats G3AINDEP G3ASCIEN G3ACLEAR G3ASKILL G3ASECUR G3ASTATU
G3AOWNID G3ACLIMA G3ALEARN G3AINCOM G3AINFLU G3ACHALL G3ACAREE
G3ALEISU G3AMANAG G3AWTEAM G3AUSEFU G3AVARIE G3AFAMIL (F2.0).
Desc G3AINDEP G3ASCIEN G3ACLEAR G3ASKILL G3ASECUR G3ASTATU
G3AOWNID G3ACLIMA G3ALEARN G3AINCOM G3AINFLU G3ACHALL G3ACAREE
G3ALEISU G3AMANAG G3AWTEAM G3AUSEFU G3AVARIE G3AFAMIL.

*FACTOR ANALYSIS.
FACTOR
/VARIABLES G3AINDEP G3ASCIEN G3ACLEAR G3ASKILL G3ASECUR
 G3ASTATU G3AOWNID G3ACLIMA G3ALEARN G3AINCOM G3AINFLU
 G3ACHALL  G3ACAREE G3ALEISU G3AMANAG G3AWTEAM G3AUSEFU
 G3AVARIE G3AFAMIL
/MISSING PAIRWISE
/ANALYSIS G3AINDEP G3ASCIEN G3ACLEAR G3ASKILL G3ASECUR
 G3ASTATU G3AOWNID G3ACLIMA G3ALEARN G3AINCOM G3AINFLU
 G3ACHALL G3ACAREE G3ALEISU G3AMANAG G3AWTEAM G3AUSEFU
 G3AVARIE G3AFAMIL
/PRINT UNIVARIATE INITIAL CORRELATION SIG DET KMO INV REPR AIC
 EXTRACTION ROTATION FSCORE
```

```
/FORMAT SORT BLANK(.40)
/PLOT EIGEN ROTATION
/CRITERIA MINEIGEN(1) ITERATE(25)
/EXTRACTION PC
/CRITERIA ITERATE(25)
/ROTATION VARIMAX
/SAVE AR(ALL)
/METHOD=CORRELATION .

*FACTOR ANALYSIS WITHOUT VARIABLE G3ACLEAR.
FACTOR
/VARIABLES G3AINDEP G3ASCIEN G3ASKILL G3ASECUR G3ASTATU
 G3AOWNID G3ACLIMA G3ALEARN G3AINCOM G3AINFLU G3ACHALL
 G3ACAREE G3ALEISU G3AMANAG G3AWTEAM G3AUSEFU G3AVARIE
 G3AFAMIL
/MISSING PAIRWISE
/ANALYSIS G3AINDEP G3ASCIEN G3ASKILL G3ASECUR G3ASTATU
 G3AOWNID G3ACLIMA G3ALEARN G3AINCOM G3AINFLU G3ACHALL
 G3ACAREE G3ALEISU G3AMANAG G3AWTEAM
 G3AUSEFU G3AVARIE G3AFAMIL
/PRINT UNIVARIATE INITIAL CORRELATION SIG DET KMO INV REPR
 AIC EXTRACTION ROTATION FSCORE
/FORMAT SORT BLANK(.40)
/PLOT EIGEN ROTATION
/CRITERIA MINEIGEN(1) ITERATE(25)
/EXTRACTION PC
/CRITERIA ITERATE(25)
/ROTATION VARIMAX
/SAVE AR(ALL)
/METHOD=CORRELATION .

*RENAMING FACTOR VARIABLES.
RENAME VARIABLES
 (FAC1_6  = AUTO )
 (FAC2_6  = INSTAT)
 (FAC3_6  = SOCOPT)
 (FAC4_6  = TASK)
 (FAC5_6  = LEARN).

*LABELLING FACTOR VARIABLES.
VAR LAB
  AUTO   "Autonomy (comp)"
  INSTAT "Income and status (comp)"
  SOCOPT "Social opportunities (comp)"
  TASK   "Task Significance(comp)"
  LEARN  "Learning opportunities(comp)".

*SAVE DATASET.
SAVE OUTFILE= 'd:\my
document\mahe_modules\13modul7_2_finished\grad_mahe10.sav'.
```