Asking PISA-Questions to Bachelor students

ABSCHLUSSARBEIT

zur Erlangung des akademischen Grades
Bachelor of Science
(B.Sc.)
im Fach Statistik

eingereicht an der
Wirtschaftwissenschaftlichen Fakultät
Humboldt-Universität zu Berlin

von
Frau Vinh Hanh Lieu
geboren am 03.11.1979 in Berlin

Gutacter:
Prof. Dr. Wolfgang Härdle

Betreuer:
Dr. Sigbert Klinke

eingereicht am: 17. Dezember 2007

Asking PISA-Questions to Bachelor students

A Bachelor Thesis Presented

by

Vinh Hanh Lieu (500323)

Tester: **Prof. Dr. Wolfgang Härdle**

Director: **Dr. Sigbert Klinke**

Humboldt University, Berlin

Economics Faculty

Institute of Statistics and Econometrics

Berlin, December 17, 2007

Declaration of Authorship

I hereby confirm that I have authored this bachelor thesis independently and without use of others than the indicated resources. All passages, which are literally or in general matter taken out of publications or other resources, are marked as such.

Vinh Hanh Lieu

Berlin, December 17, 2007

Abstract

The Programme for International Student Assessment (PISA) is an internationally standardized assessment to 15-year-old pupils in schools. With the same purpose, the questionnaire made by Dr. Duller (Austria) was conducted at the university Humboldt, Berlin. We would like to assess how good students are at mathematics and statistics. Which part is the most difficult problem for students? What can we do to improve it? Factor analysis will be done in SPSS and Mplus to consider relationships between questions.

Keywords:
PISA Study, Factor analysis, tetrachoric correlation, Questionnaire

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Why did I name my thesis "Asking PISA-Questions to Bachelor students"?
Does it have something relevant to The Programme for International Student Assessment (PISA)?
OECD (Organisation for Economic Co-operation and Development) began work on PISA in the mid-1990s. PISA is an internationally standardised assessment that was jointly developed by participating countries and administered to 15-year-olds in schools. Every test consists of four parts. They are problem soving, reading, mathematical and scientific literacy. PISA assesses how well students can apply the knowledge and skills they have learned at school to real-life challenges.
The survey was implemented in 43 countries in the 1st assessment in 2000, in 41 countries in the 2nd assessment in 2003, in 57 countries in the 3rd assessment in 2006 and 62 countries have signed up to participate in the 4th assessment in 2009. More information about PISA can be found hier.
Dr. Duller and her colleagues at the university Linz, Austria made a questionnaire to evaluate mathematical and statistical comprehension of students in their university. With the same concept she named this study a kind of PISA-Survey. This survey was conducted three times (2001, 2003, 2004) in Austria (Duller C., A kind of PISA-survey at university, COMPSTAT 2004).
This is the first time we have done this test at the university Humboldt, Berlin. We also would like to assess how good students are at mathematics and statistics. Which part is the most difficult problem for them?

The thesis is organized as follows. The questionnaire and the data structure will be introduced in the chapter 2. Some applied statistical methods will be shown in the chapter 3. In Chapter 4 the results of the survey according to some categorized variables (age, gender, semester, kind of study administration, study course, directly going to university) are presented. Chapter 5 focuses on presenting exploratory and confirmatory factor analysis of binary variables with statistical software SPSS and Mplus. Compare the results of Humboldt University with these of Linz University, Austria is the central topic of the chapter 6. In the last part of this thesis 7 some conclusions will be considered.

Chapter 2

Data

2.1 Questionnaire

The questionnaire in Figure 2.1 consists of twelve questions. It was taken from Dr. Duller at the university Linz, Austria. One of them is literary question. The other eleven questions deal with basic mathematics or statistics. Most of them are multiple choice with six different possibilities of response. The topics were calculation (without calculator) of percentages and fractions, interpretation of graphics and estimation of square roots and percentages.

2.2 Data structure

The test was done in three classes Statistics 1, Introduction to Econometrics, and Macro econometrics 2. The class Statistics 1 is for the second-semester students. The two other classes are for the fourth-semester students. The samples were not selected randomly. The students were chosen, because they took part in basic courses in statistics (at least the class Statistics 1).
158 respondents were asked. 85 students in the class Statistics 1 have completed the questionnaire. In the class Introduction to Econometrics 35 students have taken part in the test. The rest was 38 students in the lecture Macro econometrics 2.
There are a total of six variables about personal data. These are Age, Gender, Study course, Semester, Study administration, Directly going to university. 84 of them are female. The variable Age 2.2 was categorized, which makes some later analyzes easier and clearer. After categorizing we have 26 respondents at the age from 19-20, 61 respondents at the age from 21-22, 31 respondents at the age from 23-24, 40 respondents over 25. The variable Age was categorized like this because the difference of the knowledge between 19-year-olds and 20-year-olds would be not big. The study course was divided into three kinds. Betriebswirtschaftslehre (BWL) students were 71. Volkswirtschaftslehre (VWL) students were 60. 27 students were Wirtschaft pedagogy, Lehramt Wirtschaftswissenschaft and other students 2.3.

Humboldt-Universität zu Berlin Wirtschaftswissenschaftliche Fakultät Institut Statistik und Ökonometrie	Juli 2007

Persönliche Daten:

Alter ☐

Geschlecht ☐ weiblich ☐ männlich

Studiengang ☐ BWL Dipl ☐ VWL Dipl ☐ Wipäd Dipl
☐ BWL BA ☐ VWL BA ☐ LA WiWi
☐ BWL NF/ZF ☐ VWL NF/ZF ☐ Sonstige

In welchem Semester befinden Sie sich? ☐ Semester

Art der Studienzulassung ☐ mit deutschen Abitur
☐ EU-Bürger(in) mit hochschulqualifizierenden Abschluss
☐ Nicht-EU-Bürger mit hochschulqualifizierenden Abschluss
☐ Sonstige

Haben Sie direkt nach dem Abitur (oder entsprechenden Schulabschluss) mit dem Studium begonnen?
☐ Ja ☐ Nein

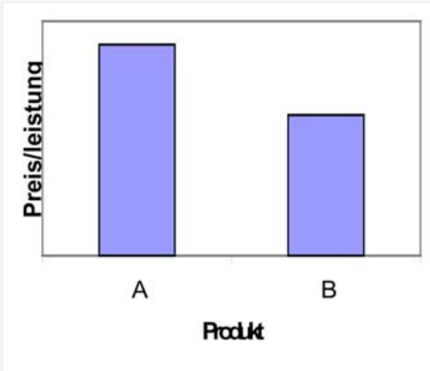
1. Welcher der folgenden Schriftsteller schrieb den „Faust“?
   ☐ Schiller ☐ Shakespeare ☐ Goethe
   ☐ Molière ☐ Kleist ☐ Hölderlin

Verwenden Sie bitte keinen Taschenrechner !!!

2. Wie viel Prozent sind 30% von 70%?
   ☐ 3% ☐ 17% ☐ 21%
   ☐ 30% ☐ 37% ☐ 70%
3. Der Bruch 1/40 kann auch ausgedrückt werden durch
   ☐ 0,40 ☐ 4/100 ☐ 0,25
   ☐ 0,040 ☐ 1/25 ☐ 0,025
4. 40% bedeutet soviel wie
   ☐ jeder vierte ☐ einer von 40 ☐ vier von zehn
   ☐ ein Viertel ☐ 1/25 ☐ zehn von 400
5. Eine Milliarde wird dargestellt durch
   ☐ 1000000 ☐ 10000000 ☐ 100000000
   ☐ 1000000000 ☐ 10000000000 ☐ 100000000000

Figure 2.1: Questionnaire

6. Untenstehendes Diagramm stellt den Preis-Leistungs-Index von zwei Produkten dar. Für welches Produkt entscheiden Sie sich?



☐ A ☐ B

7. An welcher Stelle steht bei folgender Multiplikation der Dezimalstrich?
0,000264 * 0,0074 =

☐ 0,000019536 ☐ 0,0000019536
☐ 0,00000019536 ☐ 0,000000019536
☐ 0,0000000019536 ☐ 0,00000000019536

8. Die Wurzel aus 0,5 ist

☐ größer als 0,5 ☐ gleich 0,5 ☐ kleiner als 0,5

9. Wurzel aus 14641

☐ 11 ☐ 71 ☐ 121
☐ 235 ☐ 550 ☐ 739

10. 12% von 236875

☐ 860 ☐ 1465 ☐ 2435
☐ 12416 ☐ 28425 ☐ 45650

11. Ein Kilogramm sind wie viel Gramm?

☐ 10 Gramm ☐ 100 Gramm
☐ 1000 Gramm ☐ 10000 Gramm

Vielen Dank für Ihre Mitarbeit!

12. Welches der folgenden Diagramme zeigt den stärkeren Anstieg?

☐ das erste ☐ das zweite ☐ beide sind gleich

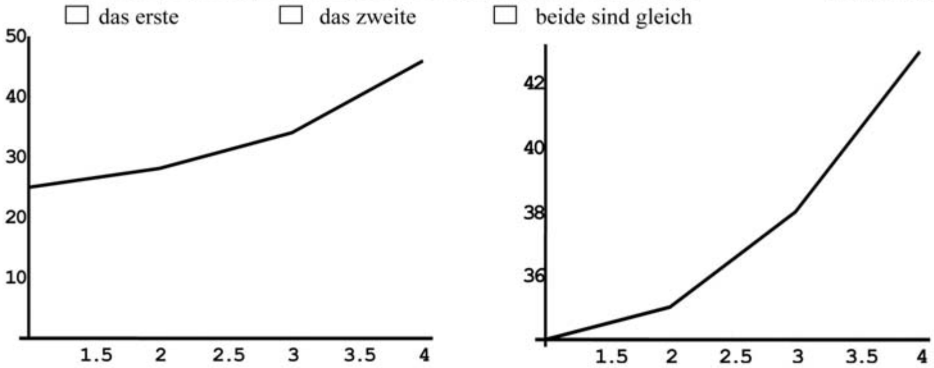


Figure 2.2: Questionnaire

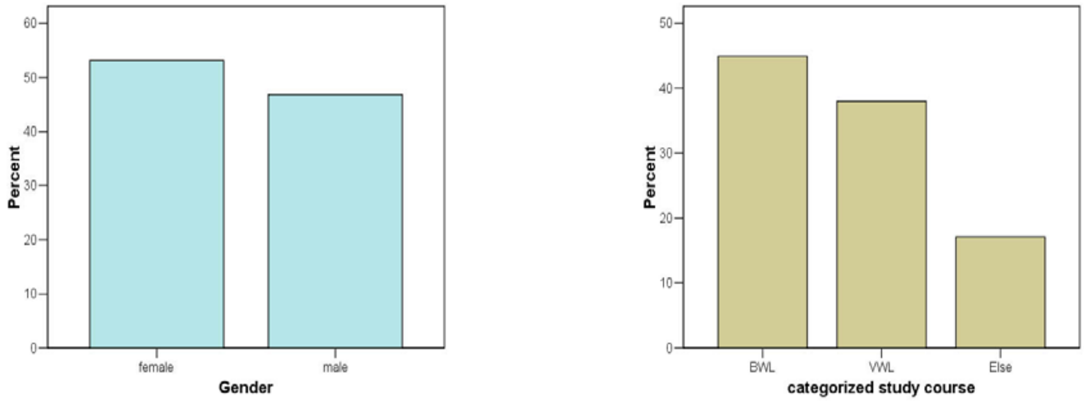


Figure 2.3: Gender and Study course

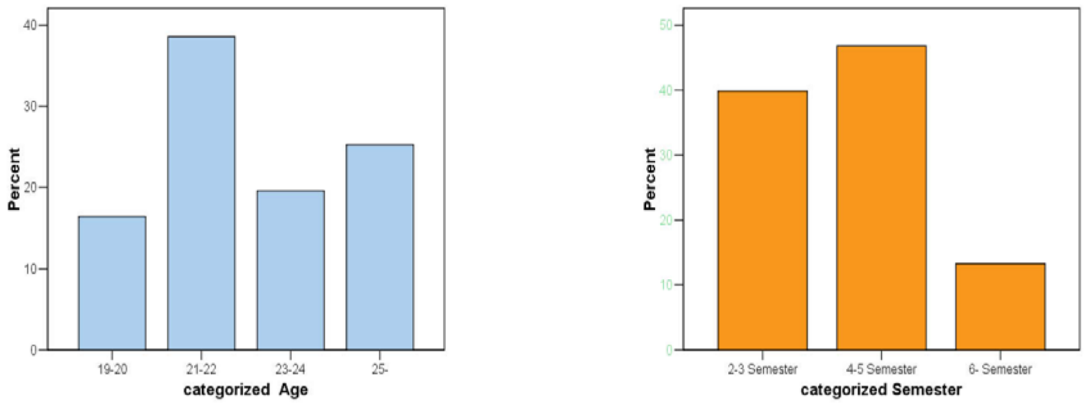


Figure 2.4: Age and Semester

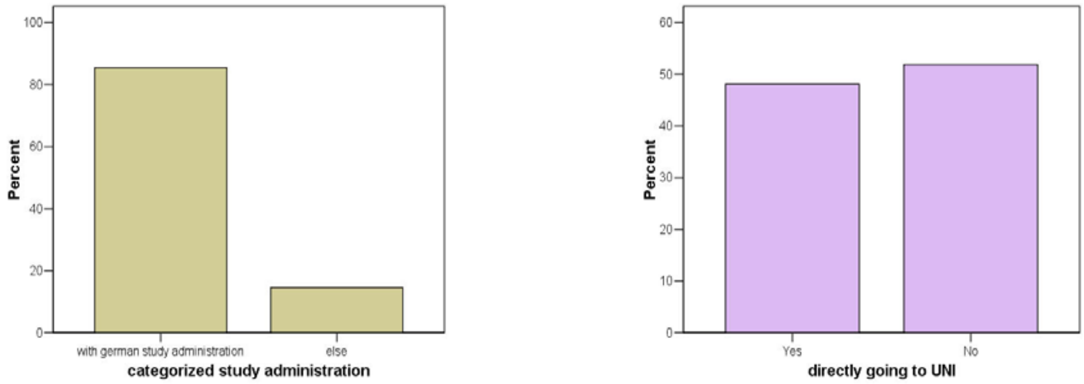


Figure 2.5: Study administration and directly going to university

63 students were in second-third semester. In fourth-fifth semester there were 74 students. From sixth semester up there were 21 students 2.4. The matriculation at the university in BWL and VWL takes place every term. That is the reason why I put the third-semester students in the group second-third semester. 135 respondents went to university with german university entrance administration (Abitur). The others were from European union, non-European union and else. The last variable is going directly to university. Less than a half went directly to university after school.

Chapter 3

Applied statistical methods

3.1 Test statistics of Mann and Whitney Test

Mann and Whitney Test (Mann-Whitney-Wilcoxon Test, Wilcoxon rank-sum test) is a non-parametric test for assessing whether two samples of observations come from the same distribution. The null hypothesis is that the two samples are drawn from a single population, and therefore that their probability distributions are equal. It requires the two samples to be independent (`http://en.wikipedia.org/wiki/Mann-Whitney_U`). The test was proposed for equal sample sizes initially by Wilcoxon (1945). In 1947 Mann and Whitney extended it to arbitrary sample sizes. The U test is calculated based on the rank over the combined samples as follows

1. Rank all the observations without regard to which sample they are in.

2. Add up the ranks in sample 1 (R_1). The sum of ranks in sample 2 (R_2) is computed with the formula $\frac{N(N+1)}{2}$ - R_1 where N is the total number of observations.

3. U-Test is given by

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \tag{3.1}$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 \tag{3.2}$$

where n_1, n_2 the sample size of sample 1, 2

$$U = min(U_1, U_2) \tag{3.3}$$

For samples above about 20 there is a good approximation using the normal distribution.

$$Z = \frac{U - m_U}{\sigma_U} \tag{3.4}$$

where

$$m_U = \frac{n_1 n_2}{2}$$

$$\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

- Z is a standard normal deviate whose significance can be checked in table of the normal distribution.
- m_U is the mean of U if the null hypothesis is true.
- σ_U is the standard deviation of U if the null hypothesis is true.

3.2 Test statistics H-Test Krustal and Wallis

H-Test Krustal and Wallis is an extension of Mann-Whitney-Wilcoxon Test, used for more than two independent samples. It is also based on the common rank over all samples (Wikipedia).

$$H = \frac{1}{B}\left[\frac{12}{n(n+1)}\sum_{p=1}^{k}\frac{1}{n_p}R_p^2 - 3(n+1)\right] \tag{3.5}$$

where

$$B = 1 - \frac{1}{n^3 - n}\sum_t (a_t^3 - a_t)$$

- R_p is the rank of the sample p
- n is the total number of observations
- n_p is the number of observations of sample p
- a_t is the number of the same value x_{ip}
- and t is index of different identical values

Compare the test statistic with the critical value of χ^2-distribution (if $n_p > 5$, for all p) with df=k-1. In the case of $n_p \leq 5$ the test statistic is H-distributed.

3.3 Reliability analysis

Reliability is the correlation between the observed variable and the true score when the variable is an inexact or imprecise indicator of the true score (Cohen and Cohen, 1983). Inexact measures may come from guessing, differential perception, recording errors, etc. on the part of the observers.
Cronbach's Alpha is a coefficient of reliability (or consistency). It measures how well a set of items (or variables) measures a single unidimensional latent construct. When data have a multidimensional structure, Cronbach's alpha will usually be low. Cronbach's alpha is calculated with the formula

$$\alpha = \left(\frac{m}{m-1}\right)\left(1 - \frac{\sum_{j=1}^{m} S_j^2}{S_Y^2}\right) \tag{3.6}$$

where

$$S_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_{ij} - \overline{X}_j)^2$$

$$\overline{Y} = \sum_{j=1}^{m} \overline{X}_j$$

$$S_Y^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(\sum_{j=1}^{m} X_{ij} - \overline{Y} \right)^2$$

- n is the number of observations
- m is the number of items in scale
- S_j^2 is the variance of item j
- S_Y^2 is the variance of total score

The standardized Cronbach's alpha can be written as a function of the number of items (m) and the average inter-correlation ($\overline{r}$) among the items.

$$\alpha = \frac{m\overline{r}}{1 + (m-1)\overline{r}} \tag{3.7}$$

From this formula one can see that if one increases the number of items, Cronbach's alpha will rise. Additionally, if the average inter-item correlation is low, alpha will be low and reversed. The range of the alpha is from 0 to 1. A reliability coefficient of 0,70 or higher is considered "acceptable" in most research situations.

3.4 Exploratory factor analysis in SPSS

Factor analysis is a statistical data reduction technique used to explain variability among observed random variables in terms of fewer unobserved random variables called **factors** (Wikipedia). The observed variables are modeled as linear combinations of the factors, plus "error" terms or not. Unobserved factors are much more interesting to researchers than observed variables themselves, because they give a better understanding of customer behavior in marketing research, or of certain behavior in some other fields of economics, politic sciences, ...

Exploratory factor analysis (EFA) is done when we do not know how the model looks like. The method lets the dependencies of the variables find the model.

3.4.1 Kaiser-Meyer-Olkin measure and Bartlett's test of sphericity

The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) is an index for comparing the magnitudes of the observed correlation coefficients to the magnitudes of the partial correlation coefficients.

$$KMO = \frac{\sum\sum r_{ij}^2}{\sum\sum r_{ij}^2 + \sum\sum a_{ij}^2} \qquad \text{für } i \neq j$$

where

Value of KMO	Assessment
0,9 - 1	marvelous
0,8 - 0,9	meritorious
0,7 - 0,8	middling
0,6 - 0,7	mediocre
0,5 - 0,6	miserable
under 0,5	unacceptable

Table 3.1: Assessment of KMO Test

- r_{ij} is the observed correlation coefficients
- a_{ij} is the corresponding partial correlation coefficients

The partial correlation between two variables is the correlation between two variables when the influence of all remaining variables is eliminated. The range of KMO is from 0 to 1. The KMO is near the value of 1, when the partial correlation coefficients are small. The small partial correlation between two variables indicates that one of both variables is maybe much correlated with the remaining variables. It makes doing the factor analysis meaningful.
Bartlett's test of sphericity is another indicator of the strength of the relationship among variables. It is used to test the null hypothesis that the variables in the population correlation matrix are uncorrelated.
When KMO is big enough (greater than 0.5) and the null hypothesis of Bartlett's test is rejected, factor analysis can be done.

3.4.2 Anti-Image matrix

The anti-image correlation matrix contains the negatives of the partial correlation coefficients. Most of the off-diagonal elements should be small in a good factor model. The diagonal elements on the anti-image correlation matrix are the KMO individual statistics for each variable. They are called Measure of Sampling Adequacy (MSA). MSA has the same meaning with KMO. MSA-value is calculated for every variable with the following formula.

$$MSA_i = \frac{\sum\limits_j r_{ij}^2}{\sum\limits_j r_{ij}^2 + \sum\limits_j a_{ij}^2} \qquad \text{für } i \neq j$$

MSA-value of variables which are smaller than 0,5 should be excluded from factor analysis (according to KMO Assessment).

3.4.3 Rotation of factors

Varimax rotation is an orthogonal rotation of the factor axes to maximize the variance of the squared loadings of a factor on all the variables in a factor matrix.

Promax rotation rotates the factor axes allowing to have an oblique angle between them.
A rotated solution yields results which make it as easy as possible to identify each variable with a single factor.

3.4.4 Factor scores

All three methods (*Regression*, *Bartlett* and *Anderson-Rubin*) give factor scores with the mean of 0. If one uses method principle components to do the factor analysis, all three methods will give the same result since the factor scores in this case are not estimated, but are computed on the basis of equation (linear combination between factor variables and observed variables). The principal components method assumes that factors are uncorrelated. It also assumes that the communality of each item sums to 1 over all factors, implying that each item has zero unique variance. Factor scores are determined by 3.10.

$$Z_{(n\times m)} = F_{(n\times Q)} A^T_{(Q\times m)} \tag{3.8}$$

$$R = AA^T \tag{3.9}$$

$$F = ZA(A^TA)^{-1} \tag{3.10}$$

where

- Z is the matrix of standardized observed values
- F is the matrix of factor scores
- A is the matrix of factor loadings
- R is the Pearson correlation matrix

3.5 Exploratory factor analysis in Mplus

In Mplus, factor analysis uses a *tetrachoric* correlation matrix for binary variables. Tetrachoric correlation is used when both variables are dichotomies which are assumed to represent underlying bivariate normal distributions. Tetrachoric correlation can be a nonpositive definite correlation matrix when one of eigenvalue is negative (reflecting violation of normality, outliers, or multi collinearity of variables).
Let y_i (i=1, ..., n) be a binary response, y_i^* be an underlying continuous variable. The formulation is closely related to the ordinary factor analysis model for quantitative variables.

$$y_{ij} = \begin{cases} 1, & \text{if } y_{ij}^* \geq \tau_j \\ 0, & \text{otherwise} \end{cases} \tag{3.11}$$

$$y_i^* = \Lambda\eta_i + \epsilon_i \tag{3.12}$$

where

- τ_j is threshold parameter j=1, ..., p
- $\Lambda_{p\times Q}$ is matrix of factor loadings
- η_i is vector of latent variables
- ϵ_i is a vector of residuals

The structural part of the model is given by

$$\eta_i = \alpha + B\eta_i + \Gamma x_i + \zeta_i \tag{3.13}$$

where

- $\alpha_{m\times 1}$ is vector of latent intercepts
- $B_{m\times m}$ is a matrix of dependent latent variable slopes with zero diagonal elements, assumed that I-B non-singular
- $\Gamma_{m\times q}$ is a matrix of covariate slopes
- x_i is a vector of observed covariates
- ζ_i is a vector of latent variable residuals

The mean vector μ_i^* and covariance matrix Σ_i^* of y_i^* are derived under three assumptions

- ϵ_i are i.i.d distributed with mean zero and diagonal covariance matrix Θ
- ζ_i are i.i.d distributed with mean zero and covariance matrix Ψ
- ϵ_i and ζ_i are uncorrelated

$$\mu_i^* = \Lambda(I-B)^{-1}\alpha + \Lambda(I-B)^{-1}\Gamma x_i \tag{3.14}$$

$$\Sigma_i^* = \Lambda(I-B)^{-1}\Psi(I-B)^{-1}\Lambda^T + \Theta \tag{3.15}$$

Let μ_{ij} denote mean of y_{ij} given x_i

$$\begin{aligned}\mu_{ij} &= E(y_{ij}|x_i) = 1\cdot P(y_{ij}=1|x_i) + 0\cdot P(y_{ij}=0|x_i) \\ &= P(y_{ij}^* > \tau_j|x_i) \\ &= \int_{\tau_j}^{\infty} f(y;\mu_{ij},\sigma_{ijj}^*)dy\end{aligned} \tag{3.16}$$

Because the variance of y_{ij}^* is not identifiable when binary data is observed. It is assumed that $\sum_i^*$ has unit diagonal elements, hence $\sigma_{ijj}^* = 1$, j=1, ..., p. It follows that

$$\begin{aligned}\mu_{ij} &= \int_{\tau_j-\mu_{ij}^*}^{\infty} \phi(z)dz \\ &= \Phi(-\tau_j + \mu_{ij}^*)\end{aligned} \tag{3.17}$$

and the conditional correlation of y_{ij} and y_{ik} given by σ_{ijk}

$$\sigma_{ijk} = E(y_{ij}y_{ik}|x_i) - \mu_{ij}\mu_{ik} \tag{3.18}$$

where

$$\begin{aligned} E(y_{ij}y_{ik}|x_i) &= 1 \cdot P(y_{ij} = 1, y_{ik} = 1|x_i) + 0 \\ &= P(y^*_{ij} > \tau_j, y^*_{ik} > \tau_k|x_i) \\ &= \int_{\tau_j - \mu^*_{ij}}^{\infty} \int_{\tau_k - \mu^*_{ik}}^{\infty} g(z_1, z_2|x_i; \sigma^*_{ijk}) dz_1 dz_2 \\ &= \Phi^*(-\tau_j + \mu^*_{ij}; -\tau_k + \mu^*_{ik}; \sigma^*_{ijk}) \end{aligned} \tag{3.19}$$

3.6 Confirmatory factor analysis in Mplus

Confirmatory factor analysis (CFA) is a statistical technique used to verify the factor structure of a set of observed variables. CFA allows researchers to test the hypothesis that a relationship between observed variables and their underlying latent constructs exists. The difference to EFA is every variable is just loaded on one factor. Error terms contain the remaining influence of variables.
The null hypothesis is that the covariance matrix of the observed variables is equal to the estimated covariance matrix.

$$H_0 : S = \Sigma(\theta)$$

where $\Sigma(\theta)$ is the estimated covariance matrix.

3.6.1 Estimation of model parameter

Muthén (1984) considered the weighted least square (WLS) fitting function as follows. An advantage of WLS-discrepancy function is that the assumption about skewness and kurtosis is not needed. Since these information are considered in the so-called asymptotic variance-covariance matrix W.

$$F_{WLS} = (s - \sigma(\theta))^T W^{-1} (s - \sigma(\theta)) \tag{3.20}$$

where

- s is vector of elements in empirical covariance matrix
- $\sigma(\theta)$ is vector of corresponding elements in estimated covariance matrix
- W is covariance matrix of variance and covariance of measured variables.

The vector σ is obtained by the multivariate regression of p-dimensional vector y on the q-dimensional vector x. The estimation of unknown parameter in this regression is carried out in two steps.
Consider as an example the case of two binary variables y_1 and y_2 regressed on x.

Univariate-response probit regression (UPR) 3.17 with log likelihood l_{ij} for individual i and variable j is computed

$$l_{ij} = y_{ij} \log P(y_{ij} = 1|x_i) + (1 - y_{ij}) \log P(y_{ij} = 0|x_i) \tag{3.21}$$

Bivariate-response probit regression (BPR) 3.19 with log likelihood l_{ijk} for individual i and variables j, k

$$\begin{aligned} l_{ijk} = y_{ij}y_{ik} \log P(y_{ij} = 1, y_{ik} = 1|x_i)+ \\ y_{ij}(1 - y_{ik}) \log P(y_{ij} = 1, y_{ik} = 0|x_i)+ \\ (1 - y_{ij})y_{ik} \log P(y_{ij} = 0, y_{ik} = 1|x_i)+ \\ (1 - y_{ij})(1 - y_{ik}) \log P(y_{ij} = 0, y_{ik} = 0|x_i) \end{aligned} \tag{3.22}$$

Solve this equation to get parameter

$$\sum_{i=1}^{n} \partial l(i)/\partial\sigma = 0 \tag{3.23}$$

- From maximum-likelihood estimates we will receive threshold parameter τ, coefficient of y (probit slopes) in UPR
- In the second step, ρ (residual covariance for y_j and y_k) are obtained in BPR holding τ and probit slopes fixed at the estimated values from the UPR.

3.6.2 Tests of model fit

In Mplus the chi-square test statistic, the Comparative-Fit-Index (CFI), the Tucker-Lewis-Index (TLI) and the Root Mean Square Error of Approximation (RMSEA) are given.
Chi-square test: the chi-square test checks the hypothesis, that the theoretical covariance matrix corresponds to the empirical covariance matrix. The test statistic is χ^2-distributed under the assumption of the null hypothesis. The null hypothesis will be rejected if the value of the test statistic is big. It is not used for the large samples because big n makes the test always significant.

$$\chi^2 = (n - 1)F(S, \Sigma) \tag{3.24}$$

where

- n is the number of observations
- F is the minimum of the discrepancy.

Comparative-Fit-Index: The test compares the tested model with the null model. In null model the correlation and covariance of the latent variables are assumed equal to 0.

$$CFI = \frac{(\chi_0^2 - df_0) - (\chi_t^2 - df_t)}{\chi_0^2 - df_0} \tag{3.25}$$

where

- χ_0^2 is χ^2-value of the null model
- df_0 is degree of freedom of the null model
- χ_t^2 is χ^2-value of the tested model
- df_t is degree of freedom of the tested model

The index lies between 0 and 1. When the index is bigger than 0,8, the model is acceptable. The value bigger than 0,95 indicates that is a good model.

Tucker-Lewis-Index (also called Non Normed Fit Index). It has the same meaning as Comparative-Fit-Index. The range of the index is from 0 to 1. When TFI is between 0,85 and 0,95 the model is acceptable. TFI > 0,95 the model fits the data very well.

$$TLI = \frac{\chi_0^2/df_0 - \chi_t^2/df_t}{\chi_0^2/df_0} - 1 \tag{3.26}$$

where

- χ_0^2 is χ^2-value of the null model
- df_0 is degree of freedom of the null model
- χ_t^2 is χ^2-value of the tested model
- df_t is degree of freedom of the tested model

The index is not influenced by the size sample. So one can use it for the data with many observations.

Root Mean Square Error of Approximation: The Root Mean Square Error of Approximation (RMSEA) is a measure for the model deviation per degree of freedom. A RMSEA < 0,05 shows an adequate model. When the value RMSEA of the tested model is smaller than 0,08, the fit of the model is acceptable. The model is unacceptable when the value is bigger than 0,1.

3.6.3 Factor scores

Factor scores is the linear combination of factor loadings and observed variables. They are sometimes computed for further analyzes.
Mplus calculates the factor loadings as probit regression coefficient with the method *weighted least square*. The interaction among factors and the error terms of variables are taken into consideration. After obtaining estimated parameter 3.6.1, put these in the formula

$$\eta_i = \alpha + B\eta_i + \Gamma x_i + \zeta_i \tag{3.27}$$

We can get factor scores.

Chapter 4

Results

The abbreviation of the questions was done, which facilitates the portray of graphics and analyzes. Figure questions are questions 6 and 12. Non-figure questions are the other nine questions. All questions mean eleven questions except literary question. Table 4.2 shows the total percent of correct answer of all questions and the results by gender. It also shows how big the different percent of correct answer between female and male. 98,7% of students answered the literary question correctly. Goethe is the author of "Faust". Dr. Duller wants to check how well students can handle with non-mathematical question. Question 12 BIGINC seems to be the most difficult question with the lowest percent of correct answer 43,7%. Questions PQIND (Which product would prefer?) and SQ05 take the second difficult level of all questions. In the third position questions SQ14K and 12o23K were found with 79,1% for both. The other six questions have the value bigger than 82,3% (DECI).

Question	Abbreviation
2. 30% of 70%	30o70
3. 1/40	1/40
4. 40%	40%
5. One billion	BILL
6. Price-quality-index	PQIND
7. Decimal point	DECI
8. Square root of 0,5	SQ05
9. Square root of 14641	SQ14K
10. 12% of 236875	12o23K
11. One kilogram	KILO
12. Bigger increase	BIGINC
Question 6, 12	Figure questions
9 other questions	Non-Figure questions
11 questions	All questions

Table 4.1: Abbreviation of the questions

Question	Total	Female	Male	Male-Female
2. 30o70	83,5	78,6	89,2	10,6
3. 1/40	86,7	82,1	91,9	9,8
4. 40%	87,3	77,4	98,7	21,3
5. BILL	89,9	89,3	90,5	1,2
6. PQIND	68,4	64,3	73	8,7
7. DECI	82,3	89,3	74,3	-15,0
8. SQ05	69,0	59,5	79,7	20,2
9. SQ14K	79,1	71,4	87,8	16,4
10.12o23K	79,1	78,6	79,7	1,1
11.KILO	97,5	95,2	100	4,8
12.BIGINC	43,7	39,3	48,7	9,4
Figure questions	33,5	29,8	37,8	8,0
Non-Figure questions	32,3	25	40,5	15,5
All questions	14,6	10,7	18,9	8,2

Table 4.2: Total percent of correct answer and the correct answer by sex

Question 11 KILO is no problem for most of students. All of male students (74 resp.) have the right solution for this question.
One third of students answered figure questions correctly. Non-figure questions have the result with 32,3%. 14,6% of students have completed the questionnaire excellently.
The total percent of correct answer was shown in Figure 4.1. Figure 4.2 shows the percent of correct answer according to sex. In ten of eleven questions males are much or little better than females. Question 7 about decimal multiplication is the only question in there females are better than males (89,3% vs. 74,3%).
In Table 4.2 the differences of correct answer between females and males are shown. In questions 40% and SQ05 males are much better than females. Are females really not so good as males in not-much-calculating and guessing questions? In another question about square root SQ14K males are also better with 16,4% more. Males are very good at square root calculation.
What about the result of question DECI? The result is a little amazing. Why can females be much better than males in answering this question? Or just because in the six possibilities of answer there are a lot of the number zero, which made male students confusing. Female students may be more careful in counting a sequence of number zero. Another good result for males in non-figure questions is the not bad difference 15,5%.

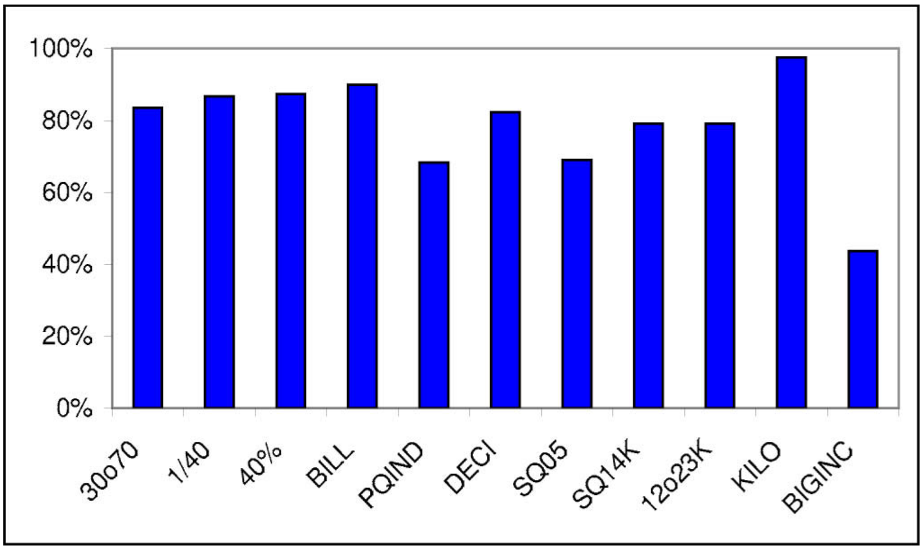


Figure 4.1: Total percent of correct answer

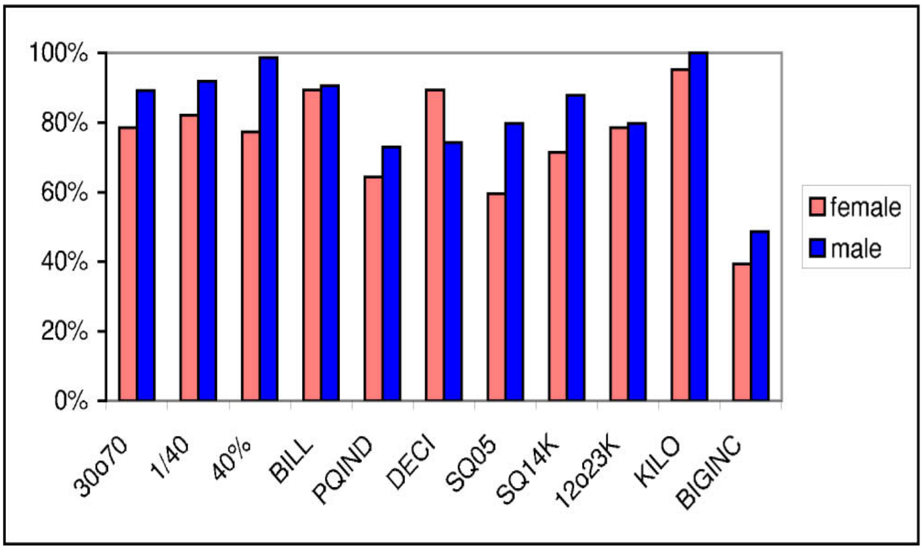
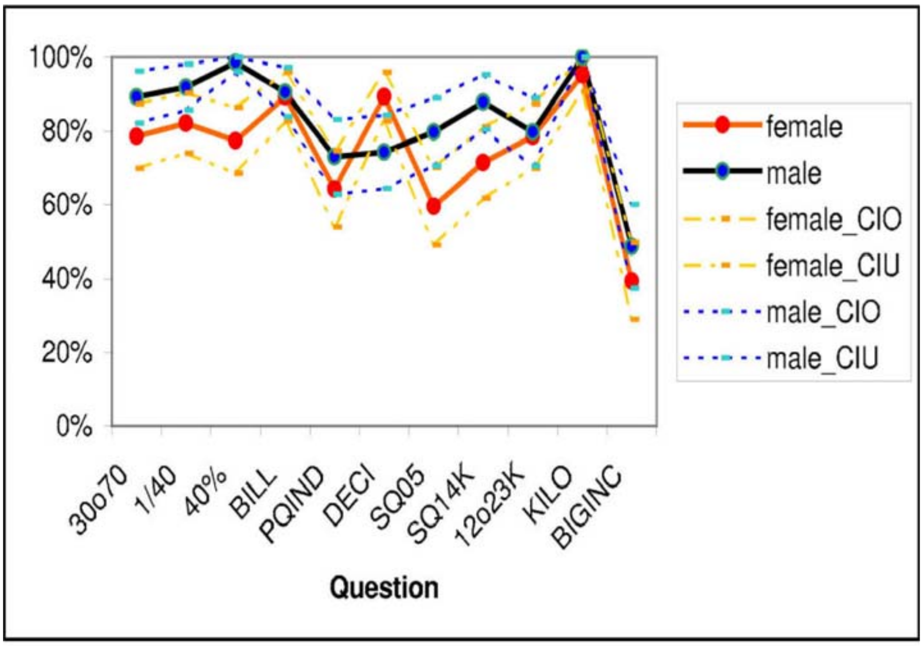


Figure 4.2: Percent of correct answer according to gender



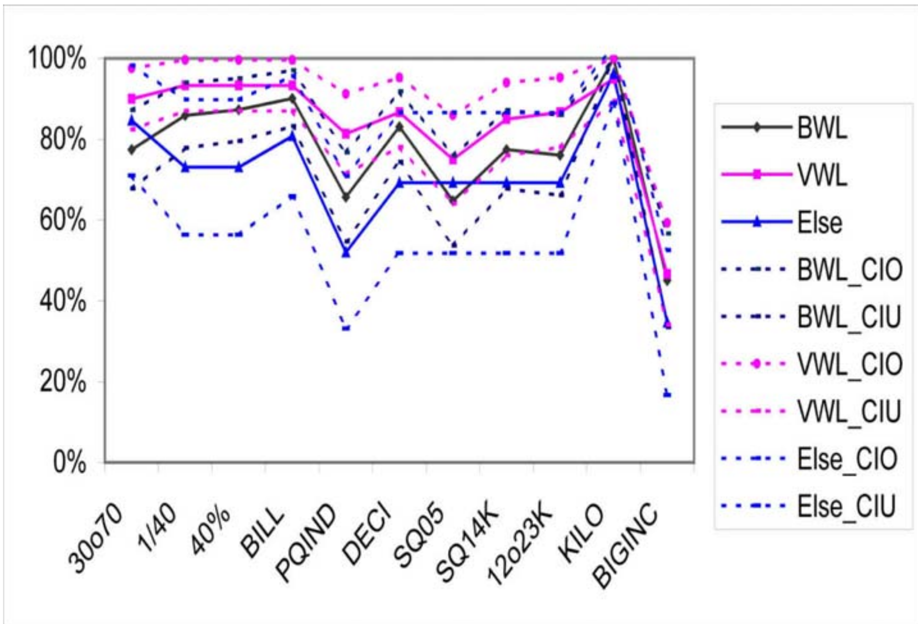
	30o70	1/40	40%	BILL	PQIND	DECI
Mann-Whitney	2778	2805	2447	3069	2838	2643
Wilcoxon	6348	6375	6017	6639	6408	5418
Z	-1,8	-1,8	-4	-,3	-1,2	-2,5
Asymp Sig.	**,07**	**,07**	**,00**	,79	,24	**,01**

	SQ05	SQ14K	12023K	KILO	BIGINC
Mann-Whitney	2480	2598	3072	2960	2817
Wilcoxon	6050	6168	6642	6530	6387
Z	-2,7	-2,5	-,2	-1,9	-1,2
Asymp Sig.	**,01**	**,01**	0,86	**,06**	,24

Figure 4.3: Confidence interval and test statistics of Mann and Whitney Test by gender

Figure 4.3 shows the percent of correct answer by gender and 5%-confidence interval (CI). Mann and Whitney Test 3.1 and H-Test of Krustal and Wallis were taken into account to check which questions are significant at which percent level. According to the test the null hypothesis *two samples female and male come from the same distribution* for four variables 40%, DECI, SQ05, SQ14K are rejected at the 5% level. The differences between women and men in three variables 30o70, 1/40, KILO are significant at the 10% level. For example Mann-Whitney U-Test of question 1/40 is 2085, which is determined from 3.3,

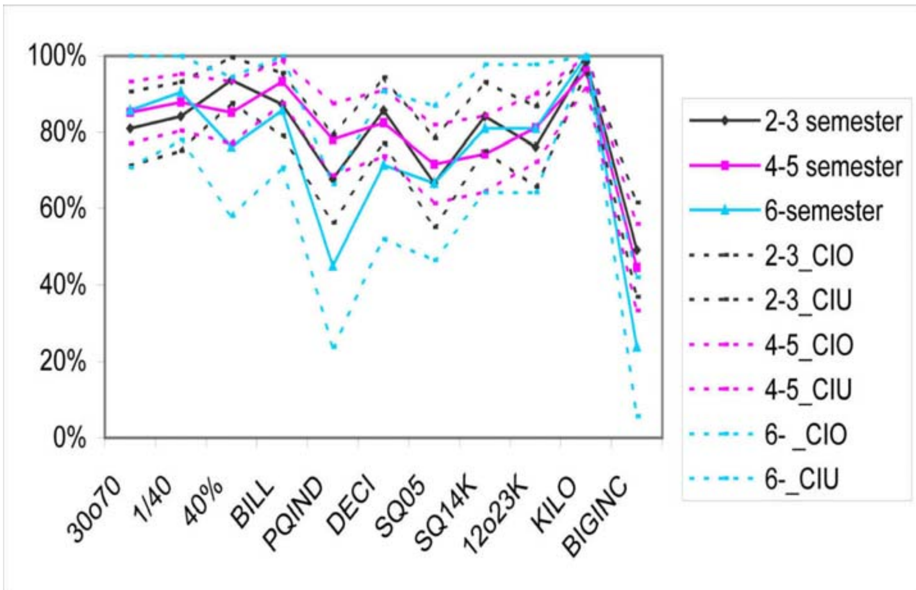
Z-value is -1,8 from 3.4. In graphics we can see that the 5%-CI of males and the 5%-CI of females are not cross each other at the questions 40% DECI, SQ05 and SQ14K. The CI of the other variables overlap somehow.
H-Test of Krustal and Wallis 3.2 was done with regard to some categorized variables study course, semester, age.
Figure 4.4 displays that VWL students are better than BWL students in almost questions. But the big differences do not exist. In only three questions 40%, 1/40, PQIND the differences are found (at the 5% level). H-Test value of 1/40 is 6,03, which is computed from 3.5 Why are VWL students somehow better than BWL students? They have maybe more mathematical and statistical courses in the module than BWL students. Graphical question PQIND (Which product would you prefer?) shows the biggest difference. Only 65,7% of BWL students answered this question correctly. The other students in the group else (Wirtschaft pedagogy, Lehramt Wirtschaftswissenschaft and other students) are not so good as the other two groups in almost questions.
Figure 4.5 shows the percent of correct answer regarding to categorized semester. There are only two significant differences at the 10% level and the 5% level respectively in questions 40% and PQIND. The fourth-semester students are not better than the second-semester students, especially in questions 40% and SQ14K even worse. The reason here is that they maybe have become used pocket calculator. Sixth- semester (up) students seem to be the worst. The result is a little amazing. So we can not conclude that the longer one is at the university, the better one is.
How is the result by age? In Figure 4.6 one also can not find clear differences among groups. CI was not done here because it will be invisible. The 21-22-year-olds are better than the 19-20-year-olds in six questions and worse in five questions. The 23-24-year-olds and 25-up-year-olds seem to be not so good as the younger.
The students going to university with german entrance study administration (Abitur) are better than the students coming from European, Non-European Union countries in almost questions. Have mathematics and basic statistics been focused in high school in Germany more than in other countries? Test statistics of Mann and Whitney Test in Figure 4.7 shows that there are significant differences at the 5% level between two groups German and Non-German in questions 40% and SQ05, at the 10% level in questions 30o70 and 12o23K.



	30o70	1/40	40%	BILL	PQIND	DECI
Chi-Square	3,76	6,03	6,21	2,87	7,53	3,43
df	2	2	2	2	2	2
Asymp Sig.	0,15	**0,05**	**0,05**	0,24	**0,02**	0,18

	SQ05	SQ14K	12o23K	KILO	BIGINC
Chi-Square	1,66	2,61	3,70	3,45	1,44
df	2	2	2	2	2
Asymp Sig.	0,44	0,27	0,16	0,18	0,49

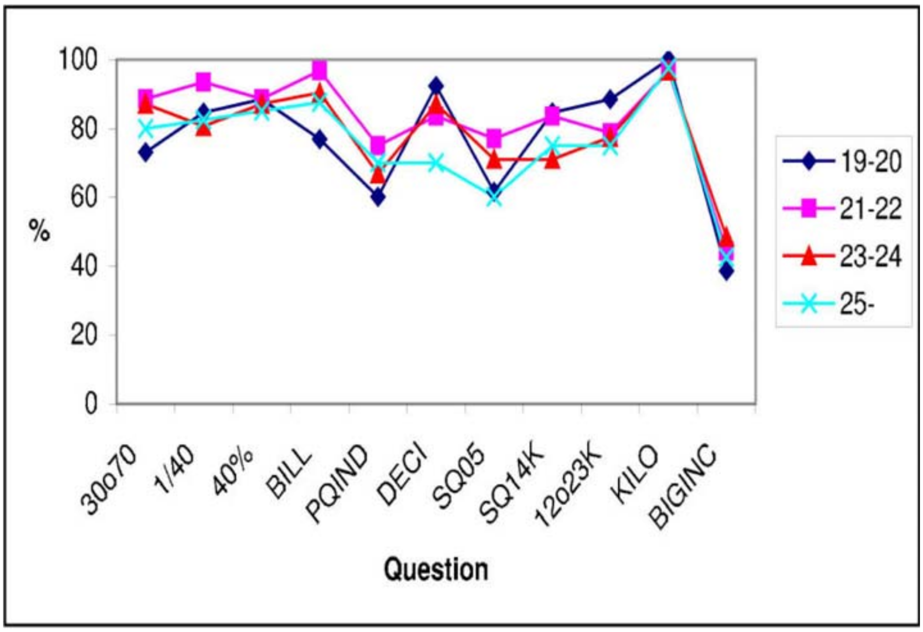
Figure 4.4: Graphics and test statistics H-Test Krustal and Wallis by study course



	30o70	1/40	40%	BILL	PQIND	DECI
Chi-Square	0,51	,70	4,93	1,77	8,91	2,19
df	2	2	2	2	2	2
Asymp Sig.	0,77	0,71	**0,09**	0,41	**0,01**	0,33

	SQ05	SQ14K	12o23K	KILO	BIGINC
Chi-Square	0,45	2,02	0,54	1,46	4,15
df	2	2	2	2	2
Asymp Sig.	0,80	0,37	0,76	0,48	0,13

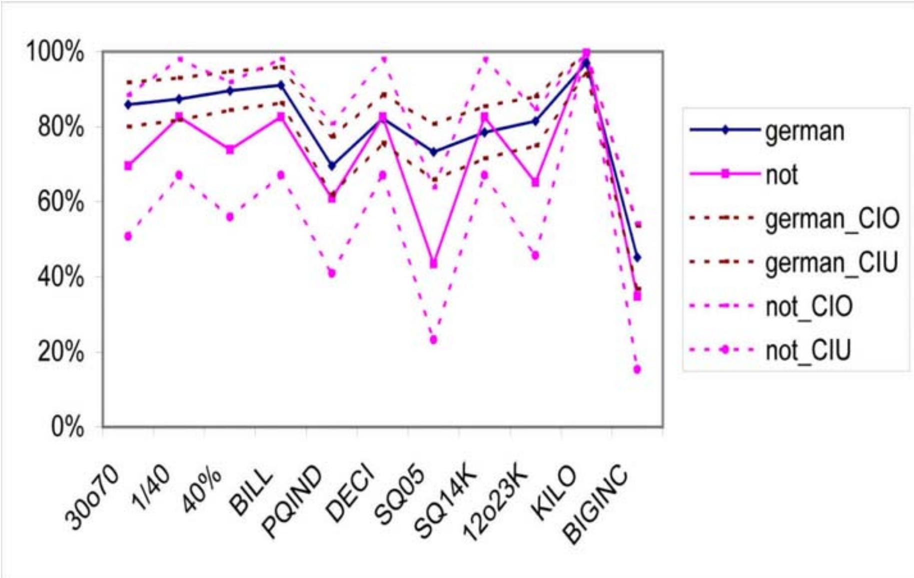
Figure 4.5: Graphics and test statistics H-Test Krustal and Wallis by semester



	30o70	1/40	40%	BILL	PQIND	DECI
Chi-Square	3,80	4,08	0,31	8,14	2,44	6,46
df	3	3	3	3	3	3
Asymp Sig.	0,28	0,25	0,96	**0,04**	0,49	**0,09**

	SQ05	SQ14K	12o23K	KILO	BIGINC
Chi-Square	4,07	2,86	1,83	0,87	0,59
df	3	3	3	3	3
Asymp Sig.	0,25	0,41	0,61	0,83	0,90

Figure 4.6: Graphics and test statistics H-Test Krustal and Wallis by age



	30o70	1/40	40%	BILL	PQIND	DECI
Mann-Whitney	1298	1478	1308	1420	1416	1546
Wilcoxon	1574	1754	1584	1696	1692	10726
Z	-1,95	-,63	-2,09	-1,25	-,83	-,05
Asymp Sig.	**0,05**	0,53	**0,04**	0,21	0,41	0,96

	SQ05	SQ14K	12o23K	KILO	BIGINC
Mann-Whitney	1089	1489	1300	1506	1391
Wilcoxon	1365	10669	1576	10686	1667
Z	-2,85	-,45	-1,77	-,83	-,93
Asymp Sig.	**0,00**	0,66	**0,08**	0,41	0,35

Figure 4.7: Graphics and test statistics by study administration

The last part of this chapter is the comparison among groups regarding to six variables gender, categorized study course, categorized semester, categorized age, german or not and directly going to university in figure, non-figure and all questions.
There are clear differences between groups with regard to sex and study course in figure questions (PQIND, BIGINC), non-figure questions and all questions. Men are better than women in three parts 4.8. Figure 4.9 portrays that VWL group is the best group. In Figure 4.10 we can find that the 6-semester students are worse than two other groups. There is no person in this group who answered all questions correctly.
The group 25-year-olds 4.11 are good in answering figure questions. Compared to other groups according to figure questions they are the best. In non-figure questions and all questions the 21-22-year-olds take the first position. Figure 4.12 said that German are much better than non-German. Not-directly going to university students are better than the other group in answering figure questions and non-figure questions 4.13. But in all questions it is a little worse.

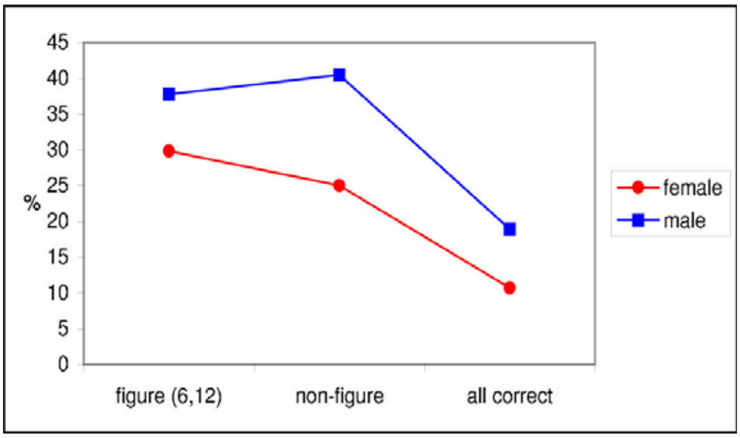


Figure 4.8: Figure, Non-figure and all questions by gender

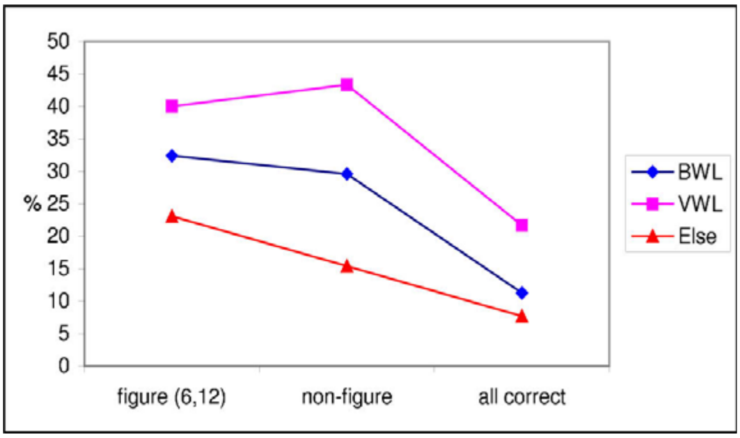


Figure 4.9: Figure, Non-figure and all questions by categorized study course

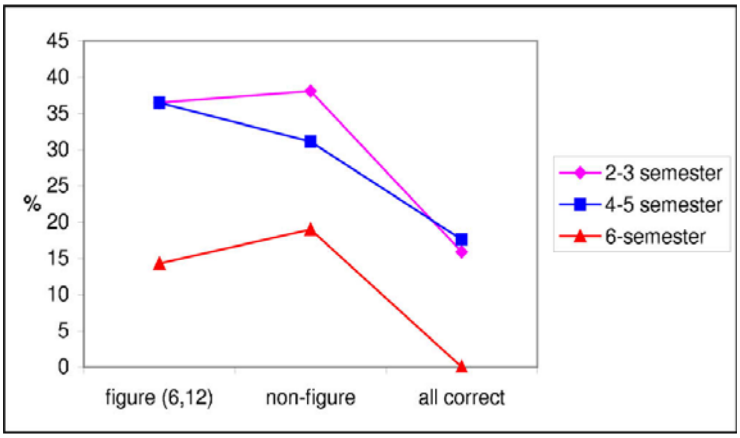


Figure 4.10: Figure, Non-figure and all questions by categorized semester

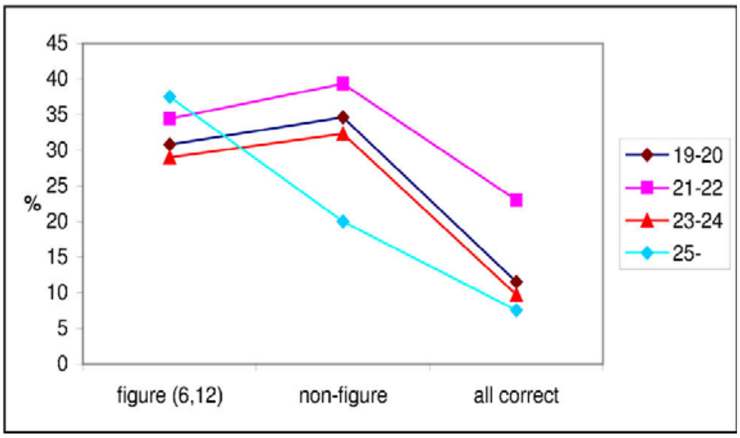


Figure 4.11: Figure, Non-figure and all questions by categorized age

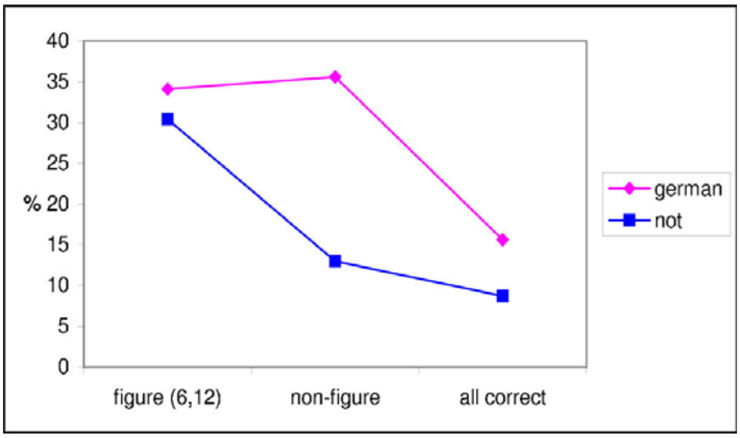


Figure 4.12: Figure, Non-figure and all questions by study administration

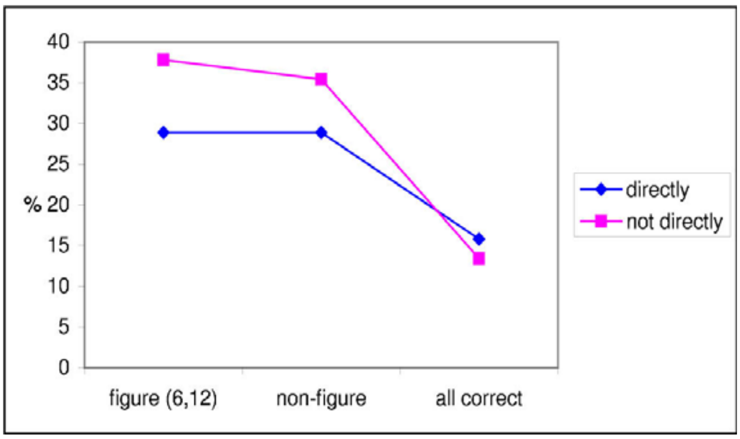


Figure 4.13: Figure, Non-figure and all questions by directly going to university

Chapter 5

Factor analysis

5.1 Factor analysis and reliability analysis with SPSS

5.1.1 Results of exploratory factor analysis

One of the assumptions of factor analysis in SPSS is that observed variables must be continuous. In the praxis people also use SPSS for binary data although the assumption is not fulfilled. I would like to introduce here some results of exploratory factor analysis for binary data in SPSS, which just gives a consideration, how SPSS reacts to the violation of the assumptions. Are the results in SPSS and Mplus different?

In Table 5.1 Measure of Sampling Adequacy 3.4.2 were found on the diagonal anti-image correlation matrix. According to Kaiser-Criteria 3.1 two questions DECI and KILO have to be excluded from factor analysis since MSA-value is smaller than 0,5. We continue to do factor analysis with nine questions. In the first step Bravais-Pearson correlation between variables is computed. In Table 5.2 we can see that nine questions are somehow correlated. Scree plot in Figure 5.1 shows three eigenvalues are bigger than one. So three factors are extracted from nine variables. In Table 5.3 about 50% of total variance explained through three factors is found.

The communalities are given in Table 5.4. The proportion of variance in each variable accounted for by the three factors is not the same. The value of communalities is between 0 and 1. Values near 0 indicate that the factors in the population do not explain the dispersion of variables at all. Values near 1 mean that the total dispersion of variables are almost explained though the factors. The value of communalities is here not high, lies from 0,35 to 0,62.

In Table 5.5 is the rotated component matrix. The relationship between the variables and the common factors is described through factor loadings. Analogous to Pearson's r, the squared factor loading is the percent of variance in that variable explained by the factor. Rotation serves to make the output more understandable and is usually necessary to facilitate the interpretation of factors.

Question	30o70	1/40	40%	BILL	PQIND	DECI	SQ05	SQ14K	12o23K	KILO	BIGINC
30o70	,56[a]	-,32	,12	,03	,05	-,08	-,15	-,08	-,14	-,28	,00
1/40	-,32	,54[a]	-,22	-,06	,04	-,09	-,12	-,13	,11	,16	,03
40%	,12	-,22	,53[a]	-,09	-,25	,11	-,01	,07	-,14	-,10	-,10
BILL	,03	-,06	-,09	,73[a]	-,02	-,05	-,04	-,10	-,04	,05	,00
PQIND	,05	,04	-,25	-,02	,60[a]	-,15	-,12	-,10	-,01	,02	-,11
DECI	-,08	-,09	,11	-,05	-,15	**,42**[a]	,10	-,06	-,13	,09	,19
SQ05	-,15	-,12	-,01	-,04	-,12	,10	,73[a]	-,17	-,15	-,01	-,03
SQ14K	-,08	-,13	,07	-,10	-,10	-,06	-,17	,73[a]	-,06	-,00	-,10
12o23K	-,14	,11	-,14	-,04	-,01	-,13	-,15	-,06	,63[a]	-,07	-,20
KILO	-,28	,16	-,10	,05	,02	,09	-,01	-,00	-,07	**,42**[a]	,07
BIGINC	,00	,03	-,10	,00	-,11	,19	-,03	-,10	-,20	,07	,60[a]

[a] Measure of Sampling Adequacy

Table 5.1: Anti-Image correlation matrix

Question	30o70	1/40	40%	BILL	PQIND	SQ05	SQ14K	12o23K	BIGINC
30o70	1,0	**,33**	-,02	,02	-,01	**,26**	,19	,19	,01
1/40	,33	1,0	,19	,12	,05	**,22**	**,21**	,03	,01
40%	-,02	,19	1,0	,13	**,27**	,12	,04	,18	,18
BILL	,02	,12	,13	1,0	,09	,09	,14	,09	,04
PQIND	-,01	,05	,27	,09	1,0	,16	,15	,12	,16
SQ05	,26	,22	,12	,09	,16	1,0	,26	,23	,12
SQ14K	,19	,21	,04	,14	,15	,26	1,0	,16	,14
12o23K	,19	,03	,18	,09	,12	,23	,16	1,0	**,23**
BIGINC	,01	,01	,18	,04	,16	,12	,14	,23	1,0

Table 5.2: Correlation matrix

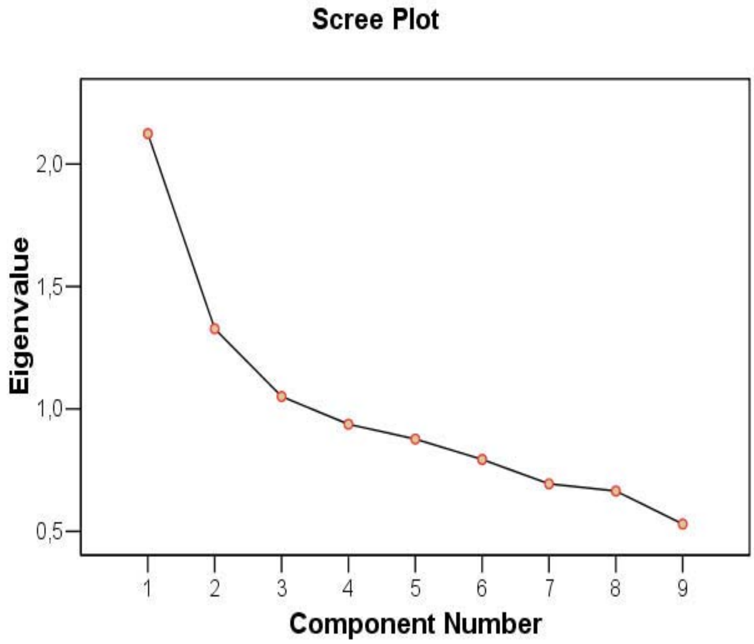


Figure 5.1: Scree plot

Question	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,12	23,60	23,60	2,12	23,60	23,60
2	1,33	14,75	38,35	1,33	14,75	38,35
3	1,05	11,67	50,02	1,05	11,67	50,02
4	,94	10,42	60,44			
5	,88	9,75	70,19			
6	,79	8,82	79,01			
7	,69	7,71	86,72			
8	,67	7,39	94,11			
9	,53	5,90	100,00			

Table 5.3: Total Variance Explained Component

Question	Initial	Extraction
30o70	1	,617
1/40	1	,621
40%	1	,557
BILL	1	,371
PQIND	1	,458
SQ05	1	,445
SQ14K	1	,350
12o23K	1	,549
BIGINC	1	,534

Table 5.4: Communalities

Question	Component		
	1	2	3
30o70	,758		
1/40	,670		
40%		,721	
BILL		,557	
PQIND		,609	
SQ05	,589		
SQ14K	,539		
12o23K			,689
BIGINC			,712

Table 5.5: Rotated Component Matrix

Factor 1	Factor 2	Factor 3
30o70	BILL	12o23K
1/40	PQIND	BIGINC
SQ05	40%	
SQ14K		

Table 5.6: Three extracted factors

KMO and Bartlett's Test Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	,652
Bartlett's Test of Sphericity Approx. Chi-Square	114,56
df	36
Sig.	0,000

Table 5.7: KMO and Bartlett's Test

Factor score	Gender	N	Mean	Std. Deviation	Std. Error Mean
factorscore1	female	84	-,20	1,11	,12
	male	74	,22	,81	,09
factorscore2	female	84	-,20	1,19	,13
	male	74	,22	,67	,08
factorscore3	female	84	-,07	,99	,11
	male	74	,08	1,00	,12

Table 5.8: Group Statistics

Table 5.6 shows three extracted factors. Due to the content of questions I named the factors as follows

- factor 1: questions requiring calculation
- factor 2: questions not-requiring calculation
- factor 3: tricky questions

KMO value is 0,652. According to Table 3.1 Criteria of KMO, it is mediocre to do this factor analysis. The observed significance level of the Bartlett's test of sphericity is .000. It is small enough to reject the null hypothesis that the variables in the population correlation matrix are uncorrelated.

In SPSS user can choose one of three different estimate methods to get factor scores 3.4.4. Factor scores were used for the further analyzes.
Here will be checked whether there is a relationship between factor variables 3.4.4 and some categorized variables. Factor variables are continuous and the variable gender is dichotomous. So independent-samples t-test is suitable for the comparison between female and male. The column 4 of Table 5.8 shows there is a somehow clear divergence in factors 1 and 2. These values are shown on the y axis of Figure 5.2. While the difference in factor 3 is not clear. Table 5.9 said that differences between female and male regarding to factor 1 and factor 2 are significant at the 5% level.
In Figure 5.2 we can see the factor scores of male have positive value. It means they are better than female. The graphics helps to see the divergence of gender clearer, especially in factor 1 (questions requiring calculation) and factor 2 (questions not requiring calculation).

In following I would like to check whether there is a relationship between factor variables and categorized study course. One-way ANOVA was used, since

factor score	Assumption	Levene's Test		T-test for Equality of Means		
		F	Sig.	t	df	Sig
factorscore1	Equal variances assumed	7,34	,007	-2,66	156	,009
	Equal variances not assumed			-2,72	150,6	**,007**
factorscore2	Equal variances assumed	20,37	,000	-2,70	156	,008
	Equal variances not assumed			-2,79	133,6	**,006**
factorscore3	Equal variances assumed	,12	,735	-1,00	156	**,319**
	Equal variances not assumed			-1,00	153,2	,319

Table 5.9: Independent Samples Test by gender

Factor variable	Assumption	Sum of Squares	df	Mean Square	F	Sig.
factorscore1	Between Groups	4,78	2	2,39	2,44	**,091**
	Within Groups	152,22	155	,98		
	Total	157,00	157			
factorscore2	Between Groups	11,35	2	5,68	6,04	**,003**
	Within Groups	145,65	155	,94		
	Total	157,00	157			
factorscore3	Between Groups	1,21	2	,61	,60	,549
	Within Groups	155,79	155	1,01		
	Total	157,00	157			

Table 5.10: ANOVA by categorized study course

categorized study course is a nominal scaled multiple grouping variable. The test statistics in Table 5.10 shows that the difference between groups (BWL, VWL, else) regarding to factor 2 is significant at the 5% level, factor 1 at the 10% level.

The positive values indicate the better group and reversed. The higher the positive value the better they are. The reason for this is that 0 was coded for the false answer and 1 for the right answer. The third group of categorized study course in Figure 5.3 are really not good at answering not-requiring-calculation questions. Regarding to categorized semester in Figure 5.4 the 6-semester students seem to be good at requiring-calculation questions, not so good as the other two groups in not-requiring-calculation questions and tricky questions. Looking at Figure 5.5 we can see that the 21-22-year-olds are very good at requiring-calculation and not-requiring-calculation questions.

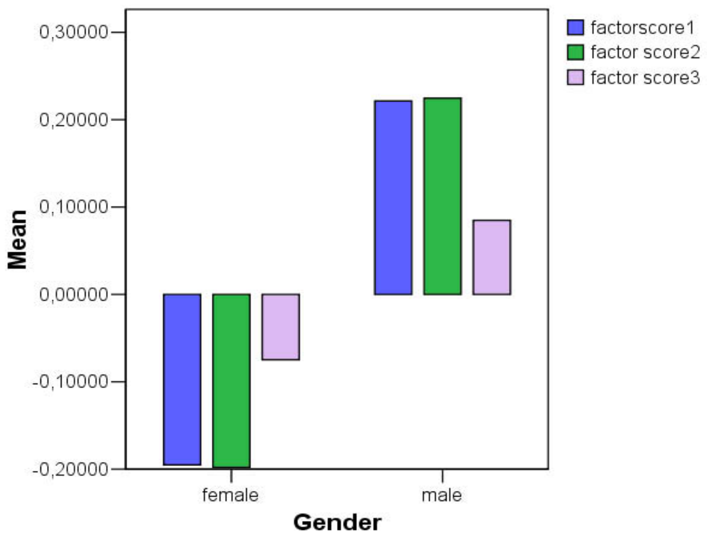


Figure 5.2: Factor score by gender

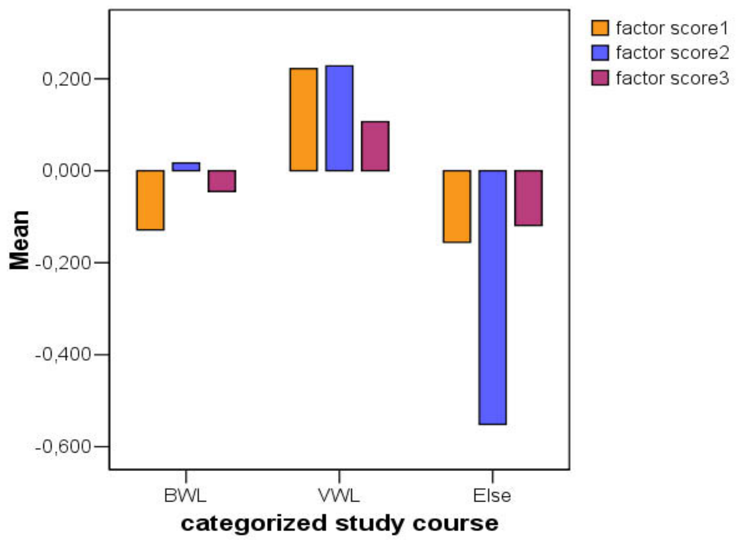


Figure 5.3: Factor score by categorized study course

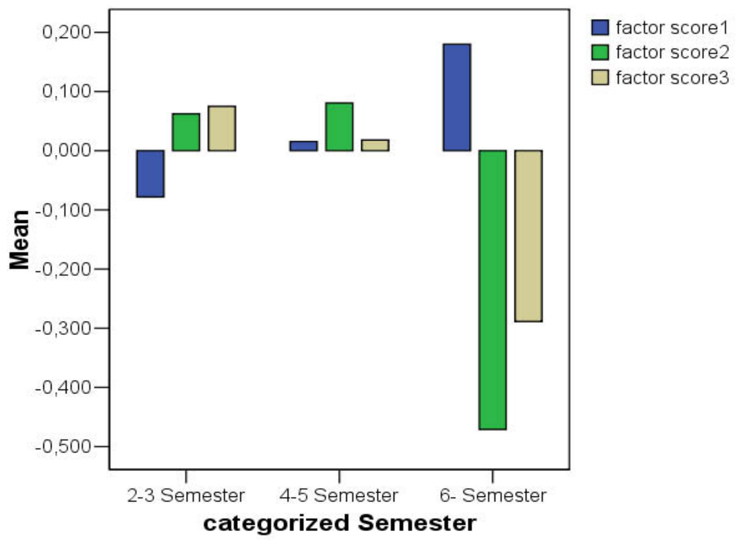


Figure 5.4: Factor score by categorized semester

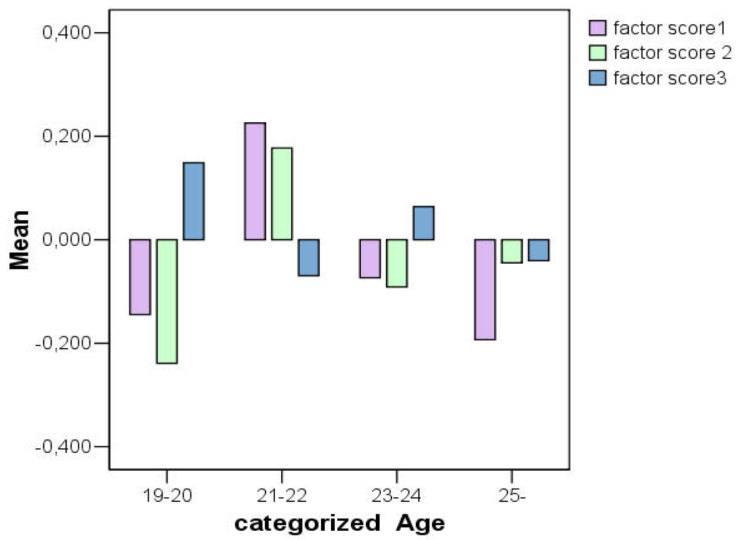


Figure 5.5: Factor score by categorized age

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
30o70	7,83	2,86	,28	,53
1/40	7,80	2,90	,28	,53
40%	7,79	2,94	,25	,53
BILL	7,77	3,07	,17	,55
PQIND	7,98	2,72	,27	,53
DECI	7,84	3,13	**,05**	,58
SQ05	7,97	2,60	,36	,50
SQ14K	7,87	2,72	,34	,51
12o23K	7,87	2,72	,34	,51
KILO	7,69	3,29	**,06**	,56
BIGINC	8,23	2,80	,18	,56

Table 5.11: Item-Total Statistics of reliability analysis

5.1.2 Reliability analysis

Table 5.11 shows that corrected item-total correlation of DECI, KILO is very small. It means that two variables are not somehow correlated with the sum of the other items. So these questions should be excluded from the analysis. Cronbach's alpha 3.6 of nine questions is 0,584. This value means these items do not have "good" reliability. Data are multi-dimensional.
From the EFA we got the three factors. Now taking the alpha of each subset of items separately. The value of Cronbach's Alpha of three factors is 0,56, 0,36 and 0,37, respectively. According to the formula of Cronbach's Alpha these values are relative small because of m.
A reliability coefficient of 0,70 or higher is considered "acceptable" in most research situations. So it is not very meaningful to do the factor analysis for these questions.

5.2 Factor analysis with Mplus

Doing exploratory factor analysis for eleven questions we got the results as follows. Four eigenvalues are bigger than one in Table 5.12. Factor 3 just loads one question DECI. According to principle of factor analysis that one factor has to load at least two variables. So in this case only three factors should be extracted.

Table 5.13 shows the three extracted factors in Mplus and SPSS.

- 1/40, SQ05, SQ14K
- PQIND, 40%
- 12023K, BIGINC

These variables are loaded together in one factor. It is the same in SPSS and Mplus. This means that these questions are somehow correlated. 59,09% of variance is explained though three factors. In Table 5.12 one can find that the

Question	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,21	29,18	29,18	3,21	29,18	29,18
2	1,80	16,36	45,54	1,80	16,36	45,54
3	1,49	13,55	59,09	1,49	13,55	59,09
4	1,03	9,35	68,44			
5	0,98	8,90	77,34			
6	0,83	7,54	84,88			
7	0,66	6,00	90,88			
8	0,53	4,81	95,69			
9	0,37	3,35	99,04			
10	0,16	1,42	100,46			
11	**-0,042**	-0,46	100,00			

Table 5.12: Total Variance Explained Component

Mplus			SPSS		
Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
30o70	40%	1/40	30o70	BILL	12o23K
KILO	PQIND	BILL	1/40	PQIND	BIGINC
	12o23K	DECI	SQ05	40%	
	BIGINC	SQ05	SQ14K		
		SQ14K			

Table 5.13: Three extracted factors in Mplus and SPSS

last eigenvalue is *negative*. That is an unusual result. However, this is often the case with tetrachoric and polychoric correlation matrices (for ordered polytomous variables). Negative definiteness of correlation matrix can be deduced from negative eigenvalues and vice versa.
After getting three factors from EFA 3.5, we continue to do CFA 3.6. This analysis can not be done completely in Mplus, since residual variance of variable 30o70 is not positive. Table 5.14 shows the residual variance of eleven questions. So factor scores can not be computed.
We could not get the results with four and three extracted factors. How is the result with two factors? Here promax rotation 3.4.3 is used to make the interpretation easier. EFA was done again to get two extracted factors. In Table 5.15 two extracted factors are shown. Only 45,54% of variance is explained through two factors. According to the content of the questions I named two factors as follows

- factor 1: not-difficult questions
- factor 2: difficult questions

Observed variable	Residual variance
30o70	**-4.78**
KILO	0.93
40%	0.53
PQIND	0.69
12o23K	0.57
BIGINC	0.74
1/40	0.57
BILL	0.87
DECI	0.98
SQ05	0.54
SQ14K	0.62

Table 5.14: Residual variance of eleven questions

Factor 1	Factor 2
30o70	40%
1/40	BILL
DECI	PQIND
KILO	SQ05
	SQ14K
	12o23K
	BIGINC

Table 5.15: Two extracted factors

Value	28.623
d.f	22
P-Value	0.1559

Table 5.16: Chi-Square Test of Model Fit

The output in Mplus gives some tests of model fit.

- Table 5.16 shows the p-value of Chi-Square Test of Model Fit is 0,15. The null hypothesis *the theoretical covariance matrix corresponds to the empirical covariance matrix* can not be rejected. Chi-Square Test of Model Fit is calculated from the formula 3.24. Since the number of observations n=158 is relative big, so the test statistic is not really convincing.
- Comparative-Fit-Index 3.25 is 0,893
- Tucker-Lewis-Index 3.26 is 0,878
- Root Mean Square Error Of Approximation (RMSEA) is 0,044

According to these critical values the model is acceptable. 89,3% of CFI means that 89,3% covariation of the data is explained through the tested model. The two factor scores with regard to some categorized variables were displayed in the following graphics. The explanation of factor scores is similar to SPSS. The graphics with semester is unusual. All of the mean factor scores of categorized semester have negative values since Mplus calculates the factor loadings as probit regression coefficient with the method *weighted least square* 3.6.3.

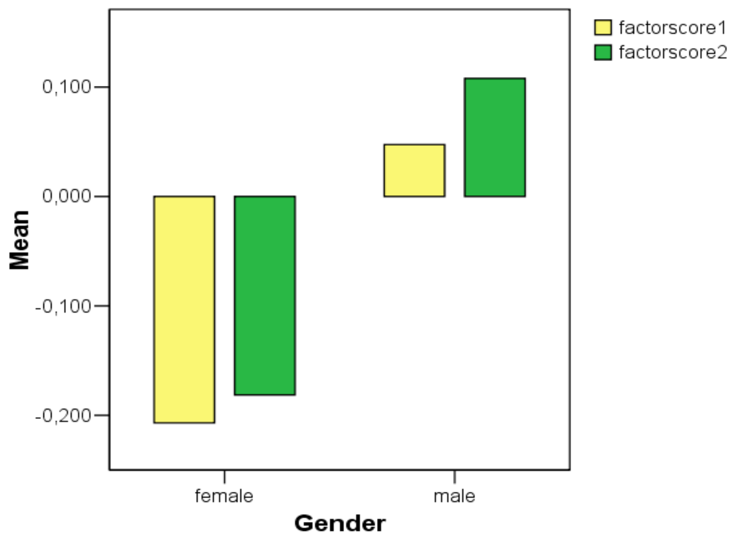


Figure 5.6: Factor score by gender

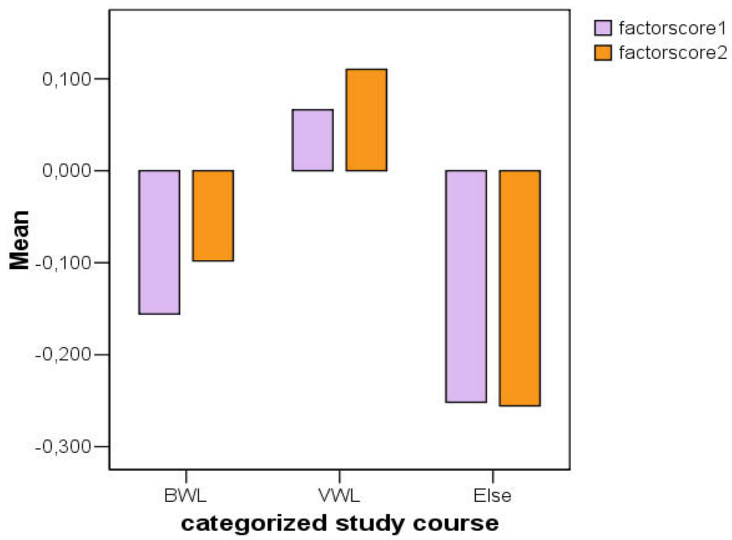


Figure 5.7: Factor score by categorized study course

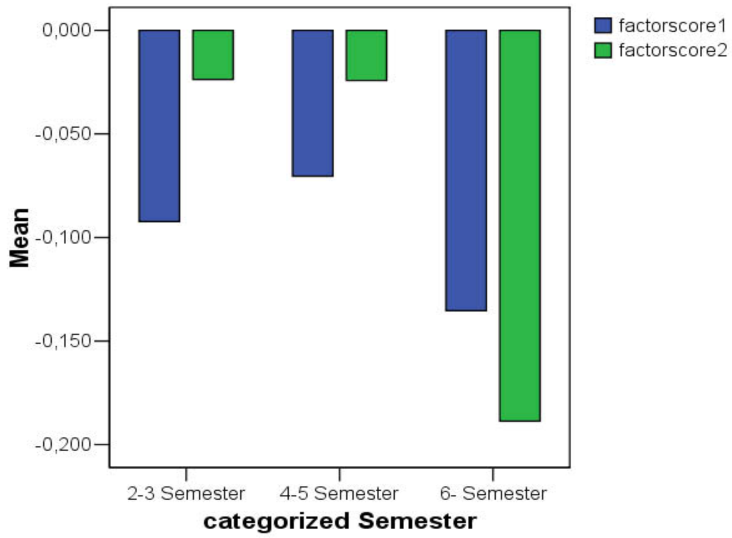


Figure 5.8: Factor score by categorized semester

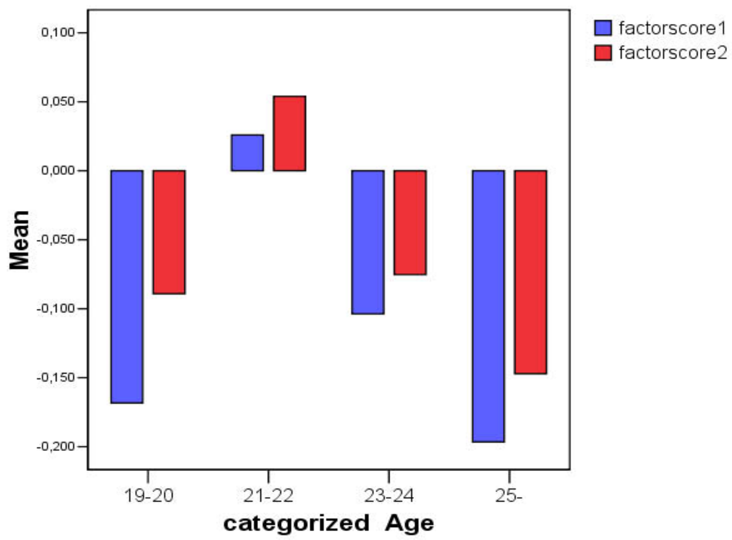


Figure 5.9: Factor score by categorized age

Chapter 6

Comparison

The last chapter of the thesis shows some comparisons between university Linz, Austria and university Humboldt (HU), Berlin Germany. Table 6.1 indicates the total result and the results with regard to gender of Linz and Humboldt. The results of Linz university are given here are the average results of three years 2001, 2003 and 2004 (COMPSTAT 2004). Linz students are as good as HU students in five questions 40%, BILL, DECI, KILO, BIGINC. Big differences between Linz and HU do not exist in these questions. In the question PQIND (Which product would you prefer?) HU students are much better than Linz students.

Linz female respondents are little better than HU ones in BIGINC (Which figure shows a bigger increase?) and question 40%. In other questions HU females are better than Linz females. How is the result of males? Like females, HU male students are also not so good as Linz students in answering question BIGINC. In question DECI HU male students are worse than Linz male students. As mentioned in the chapter 4 they are also worse than HU female. HU male

Question	Female		Male		Total	
	Linz	HU	Linz	HU	Linz	HU
PQIND	42	64	39	73	40	68
BIGINC	42	39	54	49	47	44
30o70 %	73	79	79	89	76	84
1/40	59	82	76	92	66	87
40%	82	77	92	99	86	87
BILL	85	89	93	91	88	90
DECI	86	89	82	74	84	82
SQ05	42	60	62	80	51	69
SQ14K	50	71	67	88	57	79
12o23K	63	79	76	80	68	79
Fi	18	30	22	38	20	34
Non-Fi	10	25	23	41	15	32
All	3	11	7	19	5	15

Table 6.1: Results of Linz and Humboldt

students are not really good at decimal calculation.
There are only two graphical questions 4.1 in the questionnaire. One thing is very interesting here is HU is much better than Linz in question PQIND. In the other graphical question BIGINC the result is reversed. Linz is litte better than HU relating to both female and male in BIGINC question.
The downward and upward trend of Linz and HU in Figure 6.2 and 6.3 is somehow similar. It indicates HU better or worse than Linz is almost the same with regard to gender.
The percent of students answered all questions correctly at HU is much higher than those at Linz (14,6 vs. 4,5).
An overview of the results was given. We can not conclude that Humboldt students are better than Linz students since two samples are not really comparable. The age of the youngest student at Linz was 18, at HU was 19. Additionally, the study courses of the students in two samples are different.

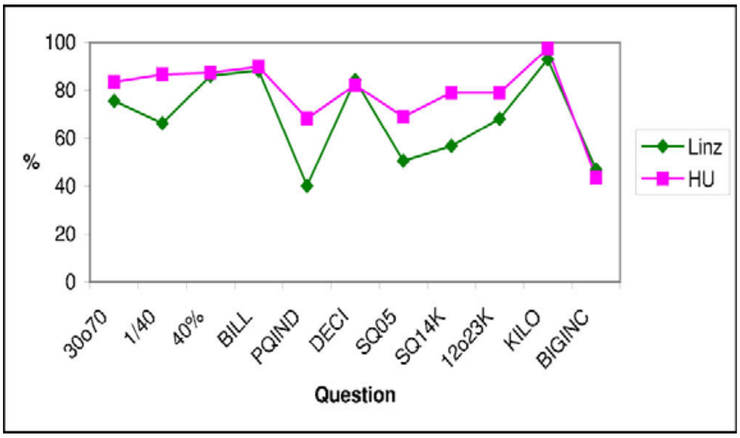


Figure 6.1: Total percent of correct answer of Linz and Humboldt University

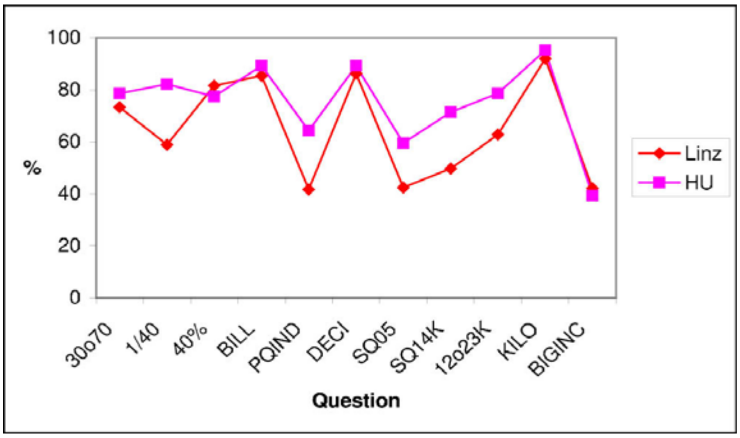


Figure 6.2: Results of Linz and Humboldt University by female

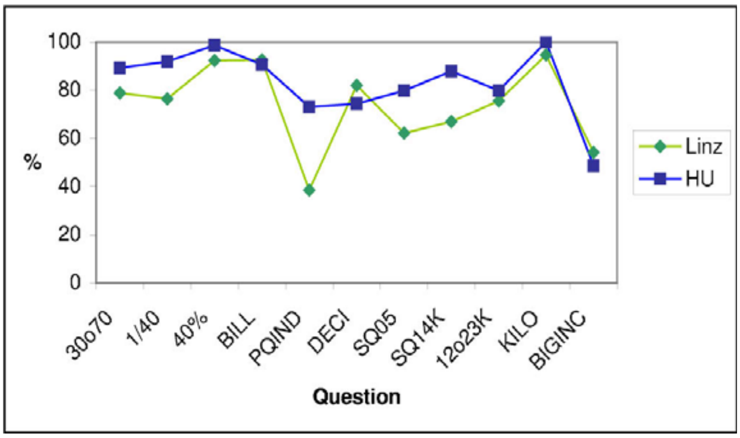


Figure 6.3: Results of Linz and Humboldt University by male

Chapter 7

Conclusion

This is the first time this PISA-survey has performed at the university Humboldt, Berlin. The results seem to be relatively good.
Students have however some problems with two graphical questions. Only 43,7% of students answered question *Which figure shows a bigger increase*? correctly. How can they assess exactly e.g. increase in prices of stocks in a certain period of time? Which share is more profitable when rate of return is given on the y axis?
The result was 68,4% for question *Which product would you prefer*?. One product is more expensive for the same quality than the other product. Students who answered this question not correctly work in importing or material purchasing department in the future, how can they make profit when they do not know which product is cheaper for the same quality?
In almost every field interpretation of diagrams or output is very important and necessary. In my opinion, more focusing on interpreting graphics will help students a lot in the future.
Some other conclusions are considered in following

- Males are better than females in almost questions except the question decimal calculation.
- VWL students seem to be the best students generally. That is predictable, since VWL students have more mathematical and statistical courses than BWL and other students in the module.
- The 21-22-year-olds found the right solution more than other groups. Most of them are in the fourth-fifth semester so they have certainly taken more classes than the younger.
- The students going to university with german entrance study administration (Abitur) are better than the students coming from European, Non-European Union countries in almost questions. Have mathematics and basic statistics been focused in high school in Germany more than in other countries?

One of the assumptions of factor analysis in SPSS is that observed variables must be continuous. In the praxis people also use SPSS for binary data although the assumption is not fulfilled. SPSS can give the result of exploratory factor analysis for binary data. Three factors were extracted from nine questions.

- factor 1: questions requiring calculation (30o70, 1/40, SQ05, SQ14K)
- factor 2: questions not-requiring calculation (BILL, PQIND, 40%)
- factor 3: tricky questions (12o23K, BIGINC)

Confirmatory factor analysis (CFA) can not be done in normal version SPSS. In order to do CFA in SPSS a macro program is required.
In Mplus one can perform exploratory and confirmatory factor analysis for binary data. The tetrachoric correlation is computed in Mplus for binary data (in SPSS Pearson correlation).
The result of exploratory factor analysis in SPSS is little similar with the result in Mplus for three factors. Because of violation of normality, outliers, or multi collinearity of variables, ... tetrachoric correlation can be a non-positive definite correlation matrix when one of eigenvalues is negative. The analysis in Mplus has given one negative residual variance so factor scores of three factors can not be calculated. So only two factors were extracted.

- factor 1: not-difficult questions (30o70, 1/40, DECI, KILO)
- factor 2: difficult questions (40%, BILL, PQIND, SQ05, SQ14K, 12o23K, BIGINC)

Confirmatory factor analysis in Mplus shows that the model of two factors is acceptable according to some critical values .

Bibliography

Brosius, F. (2006). *SPSS 14*, mitp.

Bühl, A. (2006). *SPSS 14 Einführung in die moderne Datenanalyse*, Pearson Studium.

Duller, C. (2004). A kind of PISA-Survey at university, *COMPSTAT 2004*, Physica-Verlag, pp. 975–980.

Härdle, W. and Simar, L. (2003). *Applied Multivariate Statistical Analysis*, Springer.

Jost, R. (2005). Meßmodelle und konfirmatorische faktorenanalyse.

Muthén, B. O. (1998-2004). Mplus technical appendices, *Los Angeles, CA, Muthén & Muthén* .

Muthén, Bengt O.and Toit, S. H. and Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes, *Psychometrika* .

Muthén, L. K. and Muthén, B. O. (2006). *Mplus User's Guide*.

Rönz, B. (2000). Computergestützte statistik 2.