12-07-2020

STATISTICAL MODELS FOR THE PERCEPTION OF MUSIC

Maurice Taekema

Bachelor of Science
Mathematics
dr. W.P. Krijnen
dr. M.A. Grzegorczyk

Abstract

The well-known theatre Oosterpoort in Groningen conducted a sociological research into the perception of music by conducting a questionnaire. This research tries to build a mathematical analysis for such studies of perception. By using mathematics, the reduction of dimension and the interdependencies among latent constructs is analysed. It was found that a suitable model is possible, reducing about half (62 questions) into seven factors. Then, by using the regression coefficients a structural equation model, the structure of the model was given an interpretation. Additionally, by using factor score estimation procedures, a better numerical interpretation was given to the seven factors. Item response theory was used for presenting an in-depth analysis for each item loading onto the latent factors. Then, by using the result from the latter theory, a vastly improved structural model was successfully constructed. In parallel, a software package was developed in order to improve the usability of the standard software used for these mathematical techniques.

List of Figures

Contents

1 Preface

The author would like to point out that not only mathematical models are presented in this study, but software applicable to most data sources has been made available. While using the Lavaan-package by the University of Gent in de R-language, we found out that for larger or more complex data the functions did not provide a convenient method of building structural models. Also, the estimation procedure used in the software has been improved over the last few decades. Below, we list the features of the TLV package that were used in this study.

Goal of the software. The R-code comprised in the package has a number of goals. All the goals are inspired by the same path taken as in this thesis. Therefore, the structure of the goals are somewhat sequential.

1. Handling large surveys or questionnaires can be time consuming. Therefore, the package provides a method of tidying the raw survey data. The tidying function include functions such as structuring the survey uniformly, trying to spot columns which are able to be parsed into numeric columns and detecting linear dependence - which can result in problems in estimation procedures - such that the data attains the highest quality as possible. Most importantly, however, the software provides a way of structuring the data by means of generated meta data. Information about the scale, whether a question has a reverse scale and whether its questions can be categorized.
2. Handling large surveys or questionnaires can also result in difficult, complex and foremost messy structural equation models. By means of functions such as MeasurementModels(...) and VarianceModels(...), the lavaan model is automatically generated, either based on meta data provided by the user beforehand or meta data generated in the previous step. Also, the use of a modification index - which will be discussed later - is easily implemented by means of the functions EmbedMI(lavaanmodel).
3. Factor score estimation of a structural equation model is not as straightforward as it is with a generalized linear model (by means of a predict function). The methods used in Lavaan are somewhat less diverse. Therefore, alternative methods to factor score estimation are implemented. ore estimation procedure, with the inclusion of a confidence interval.
4. For Item Response Theory, multiple sufficiently good packages have been made available for the R- programming language. Nevertheless, structuring the outputs and making an assessment methods relied on limited or non-existing functions. Therefore, the package provides a method of structuring the outputs of an item response model while also giving confidence intervals of some score estimation procedures. Also, it provides some alternatives to assessment procedures, such as the Andersen likelihood ratio test (1973), to assess both dichotomous and polytomous Rasch models.

As is a great good in science, the software has been made public and can be installed by running the code

```
# If necessary: install the package devtools by: install.packages("devtools")
library(devtools)
install_github("MauriceTaekema/TidyLatentVariables")
```

Improvements, additions and other comments about the software are very much welcomed. It should be noted that not all of the constructed functions were necessary for this study; some were constructed in order to see whether certain models or functions are possible.

A disclaimer should be provided. This package has been created in a few months and can contain problems, bugs and possibly incorrect results by accident. Please have caution when using the software. For the upcoming time, the package will probably be updated frequently, but when using the package, please try to identify whether the package has been updated recently. Due to the dependence upon other packages as well as base functions of the R-language, only maintenance of such software can provide safety in using the software. All the code used can also be found in the appendix at the end of this thesis. For each of the code, if applicable, the functions with respective inputs and outputs will be given such that it is clear for the reader what the code is about.

Maurice Taekema

2 Introduction

It is easy to lose yourself in the tangle of the beauty of mathematics. Abstract algebra, mathematical physics and other theoretical or applied fields of this formal science are with the right amount of time and a proper mindset enjoyable fields of study. However, in the field of sociology and other social sciences, mathematics can have a deterrent effect. However, mathematics is one the most versatile fields of science and applying mathematics in social domains provides a demonstration of its capabilities. One of the main branches of mathematics used outside the scope of the formal sciences, is undoubtedly statistics. Researchers within the realm of social science use statistical techniques in order to formulate and test hypotheses that relate to all kinds of (social) reality. The aim of this thesis is to analyse the use of commonly used statistical techniques in the social domain. While doing so, this study provides the necessary intuitive descriptions and exemplification of the mathematical theory.

2.1 Context of the problem

In 2016, a study was conducted at the Oosterpoort regarding visitor's perception of the performance they attended. The visitor's perception regarding the performance they attended was measured by means of a questionnaire, consisting of a vast amount of questions ($\sim$ 141). This study presents a complementary research to the already conducted study. For consistency and effectiveness, the data that has been collected will be considered for analysis. The eventual aim of this thesis is to measure the perception of the performance and overall experience of visitors of the Oosterpoort in a methodological and mathematically sound manner. At the end, the reader should be able to both formally and concretely understand the factors modeling the perception of the guests while also determine its related structure.

2.1.1 Questionnaire of Music Perception

First, we want to gain a proper understanding regarding the data. X questions have been asked to Y respondents, resulting in a vast amount of data. The survey primarily consisted of questions regarding mood, experience and the meaning and interpretation of the music performed. Some examples are: "How well do you know the music of Bach", or; "Did the music give you a sense of recognition?". In order to cope with such a large amount of data, we cluster the data. The previous research categorized the questions into two dimensions:

1. time; and
2. aggregated factors modeling the 'full' experience.

For the aggregated factors, questions related to a certain category are grouped. For instance, questions related to mood, e.g. happiness, being in a hurry, and so on are categorized as temper related. All such questions have been asked three times, such that we can track these factors over time. A glimpse and the full data can be found in the appendix. Summarising all the categories with some sample questions:

Factor	Time	Sample questions
Reasons of visiting the theatre	0	How did you travel?; Did a queue exists at arrival?
Knowledge about the composers	0	How well do you know the music of Bach?
Setting of the concert	all	Did you have a nice seat?
Meaning and Interpretation	1,2	Did you recognize the emotions in the piece?
Mood	1,2	Did you feel happy?
Experience	1,2	Did the piece move you?
Break	1,2	What activities did you conduct in the break?
Musical background	all	Do you play an instrument yourself?
Quality of the musicians	all	Are you impressed by the quality of musicians?
Final verdict	3	Was the concert worth your money?

2.2 Aim of the research

Survey data has been gathered to collect information about visitor's perception of the performance and overall experience. Constructing a survey poses a dilemma concerning the amount of questions. A limited number of questions simplifies the model and makes it easy to extract insights. However, it provides limited information and it is more difficult to prove the validity of conclusions. On the other hand, large surveys are difficult to capture in a few statistics or make even use of due to the large set of information. Moreover, with a large amount of variables it is often difficult to capture the real discriminating or essential variables with the most explanatory power. This is where a more advanced mathematical approach to survey data analysis comes into play. In order to cope with a large amount of data and arrive at methodological and mathematical sound conclusions, this research has been broken down into separable and concrete sub-problems:

1. **Dimension reduction:** Since many questions are considered for this study, reducing the dimension of the data provides both a method of preserving the ease of using the data while making minimal loss of information about the data. Each questions measures a *latent construct*, already given in the data. Now, the questions arises how well-constructed the hypothesized structure is. Therefore, the possibility of reducing the large amount of items into a fewer amount of latent constructs is an important part of this study.

2. **Quantifying and assessing factor structures** Not only should the structure be the point of study for this thesis, a numerical interpretation displaying the results of the assessed structure provides necessary information

3. **Development of a generalizable R-package** Thirdly, contributing to the scientific community always is an inherent goal of conducting research. A somewhat more separate aim for this study is the constructing of a software package in the R-language which can be used to easily reproduce the procedures of this thesis to any survey-esque data. The software automatically constructs models and provides alternative estimators.

Caution should be advised whenever dealing with actual real life data. The interpretation, the thought-process and the goal of the data should be taken into account. The next session tries to elaborate on the methods and design of the study which is believed to be suitable for the data.

2.3 Design and Methodology

The power of mathematics can be found in its inherent soundness and therefore satisfying visualisation. The Riemann Hypothesis, fractals and dynamical systems of ordinary or partial differential equations are nice demonstrations of such interpretations within this particular field. The same principle fortunately applies to statistics. Descriptive charts or graphs are well-known visualisation of survey data. However, one might question the ability to visualize hypothesis testing, or in our case the structure of musical data to measure the experience. To identify the suitable theory for this case, the dimensionality is the starting point for assessing relevant studies and papers. The known factors of the model can be considered a **latent variable**. The theory of a latent variable should be investigated before considering which statistical methods might be convenient. Within social sciences, D Kaplan (2004) and Bollen (1988) provide a valid framework for the latent construct. As multiple latent variables occur, studying the structure within these latent variables seems reasonable. Pinning down its possibilities, Hui,C., Law, K.S. and Chen Z.X. (1999) exemplified the use of structural equations to model the perceived safety behaviours in a nuclear plant. Although the context is quite distinct from the one at hand, the focal point of perception is completely alike.

Thus far, methods structuring the data have been considered. Performing statistical inferences requires a more numerical approach to the problem with the inclusion of every question. Kaplan (2004) identified the method of item response theory to be helpful in this area for handling individual responses and individuals questions, called items. Studying the practicality of the response theory, papers exemplifying the exploitation of the method provides a suitable and time-saving approach. In the light of the current time of online education due to the COVID-19 crisis, Meyer, J.P. and Zhu, S. (2013) employed an **item response theory model** for assessing and exemplifying scores of students of a test within MOOC-courses. At the stage of writing, the latter paper provides a meaningful way of identifying mathematical methods used in handling survey-esque data while also focusing on an elaborate exemplification. Nevertheless, the research focused primarily on whether a student or participant answered a question correctly. This point of view does not resemble the essence of this study, as this study investigates the agreeableness of a question placed on multiple scales. This change of thought can be considered as an extension of the theory of item response models, and is well used in multiple studies, which will be fully discussed in an upcoming section.

1. Exploratory factor analysis will be used to identify a structure and understand the data. Also, the method builds a theoretical framework for defining factors, categories and latent variables. Although exploratory factor analysis does not provide a method of assessing the hypothesized structure, it can identify, roughly, the amount of factors loading into the model and hence gives direction to dimensionality problem.

2. Then, the pre-defined structure of the data is implemented by means of Structural Equation Modeling and Path Analysis. Afterwards, the model must be assessed by means of confirmatory factor analysis and respective measures. The benefit of using a structural equation model is that the stucture, interaction and measurements are assessed by this method.

3. To give an numeric interpretation of the latent variables, estimators and Item Response Theory will be used to assess the scores of the participant

As the materials rely heavily on not well known definitions and notions, the next sections discusses each point of the methodology more elaborately.

2.3.1 Path Analysis and Measurement Models

Although mathematics was a branch of science developed by the Mesopotamian states in 3000 B.C., most statistical methods originated in the 19th century. The story for structural equation modeling is no different. Due to the substantial dependency on societal applications, similar to the respective thesis of this review, investigating the incentives of emergence of the method can be deemed beneficiary. Sewall Wrigth (1921) studied the directed dependencies of variables. It can be considered as a form of regression, with a special focus on causality. Due to the invention of computers, maximum likelihood algorithms could be developed in order to estimate the path models and therefore its causality. Although causality is not the main concern of the research, the idea of path analysis can be extended to the methods of structural equation modeling. However, due to the dimensionality of the data at hand, each factor belongs to a category. This category can be considered as an unobserved variable, which should be assessed and integrated in the mathematical framework.

By means of the addition of a measurement model, usually in the form of a partial least squares path modeling, the notion of a latent variable is introduced, solving the latter problem. Dong-Wan Ko, William P. Stewart (2002) showed how to integrate a measurement model containing attitudes. Although the paper did give a fair assessment of the model made, equivalent research should still be inspected in order to study alternative models. More recently, Tempelaar, Van der Loeff and Gijselaers (2007) analysed the relationship of students' attitude towards statistics. Setting aside the irony, the study showed a more rigid approach to evaluating the measurement model by integrating a rationale for each of the measurements, instead of merely assessing the overall model structure.

2.3.2 Covariance

Due to the causality of the models, covariance is the starting point of the model. The inclusion of factor analysis is no surprise within structural equation modeling. Musil, C.M., Jones, S.L. and Warner, C.D. (1998) showed with many others the use of factor analysis within structural equation modeling to assess the structure of the model. They discuss the connection of structural equation models and confirmatory analysis to evaluate the quality of the model. Assessing the structure is an essential prerequisite for conducting statistical inferences based on the model. Richter, N., Sinkovics, R.R., Ringle, CM, and Schlägel, (2016). argued that the covariance-based models did not mostly use best practices for developing advanced models. Frameworks lacking mathematical rigidity and unfamiliarity with the notions of linear algebra can easily result in a seemingly proper model while the right assumptions assuring the correctness are violated. Therefore, it seems only reasonable to formalize the idea of covariance while also grasping its capabilities. Fan,J., Ke, Z.T., Liu, H. and Xia, L (2014) provided theoretical insight in the theory and intuitive understanding of covariance by providing the use of the Rayleigh Quotient to the eigenvalues presented by the covariance matrix. However, the methods above are disjoint results of research focusing on specific examples and models, requiring a comprehensive text integrating all the topics listed thus far. In order to solve this, Boomsma (1999) identified the book of Bollen (1989) as the most complete and extensive material regarding structural equation models. Within the book, factor analysis, path analysis and measurement models are listed separately before the intersecting the three models. The result is a well-balanced theory providing a valid and rigid backbone for the problem given. The line of reasoning will be employed to ensure a logical footpath throughout the study. However, some technical details need to be addressed by a more extensive discussion. For this, relevant papers will be used and are described in the upcoming section.

2.3.3 Latent Variables

Assessing not merely the structure of the data nor the model, the latent variables itself need some numerical assignment. The arithmetical interpretation of a latent variable, however, is not without its controversies. Glass, Maguire (1996) generalized this position in the debate by arguing that factor scores cannot be obtained by using a mapping of weights of each score to the (latent) factor. Therefore, prudence is advised for using factor scores. T Asparouhov and B Muthén (2010) use a generalized linear model argument for assigning a value to a latent factor. However, in order to guarantee a similar notion of factor scores structure as within SEM, the methods of Bollen (1989) is used. Nevertheless, the use of estimating factor scores within Bollen cannot be considered to be fully scientifically well founded as the mathematics used is quite brief. For a elaborate view of this kind of estimation, Kari Ann Azevedo (2002) provides an assessment of the works on this issue of Bollen (1989). Due to the reviewing of the work of Bollen, the theory is quite akin to the work of Bollen ensuring a uniform and aligned account for factor estimation.

2.3.4 Item response theory

Working with latent variables and latent constructs provides a meaningful way for dealing with questionnaire data. However, information about the participants and every specific answer to a question is lost in this process. The change of point of view is nicely captured by the works of Petrillo J, Cano SJ, McLeod LD and Coon C (2015). The paper compared IRT-models to other statistical methods.By using a **Rasch model**, the model assessed the trade-off in the latent construct of the ability of a participant, and the latent construct of difficulty of a question. The difficulty can be extrapolated towards the scope of this thesis by interpreting it as the difficulty to agree with a statement, or *agreeableness of a statement*. However, classical Rasch models usually involve dichotome data. Therefore, a polytomous extension of the data must be sought. Baker, J.G., Rounds, J.B. and Zevon M.A. discuss the two main methods for a polytomous extension: the pure form of a polytomous Rasch model, the Partial credit model, and the graded response model, involving an ordinal structure of the data. Although the method seems best applicable due to the similarity in ordinal structure, the model assumes a uniform scale, whereas the partial credit model assumes no general scale. For the given data there are multiple scales, 0-1; 0-6 and 0-10; considered. This study will consider both methods as they both have unique advantages included.

For all the methods considered, the model relies on strict assumptions. All the materials considered thus far do not satisfy the rigidity of elaborating on the assumptions and necessary conditions underlying the methods. Hatzinger (2008) gives a detailed account for the constructing a proper mathematical model for item response theory. —By Andersen (1973b) a statistical test, the Andersen conditional likelihood ratio test, was developed to check all assumptions of the model simultaneously. The implementation of this theory both in the study as with software will be part of the study.

2.3.5 Integration and statistical reporting

Scientific soundness relies on the integration of theory and application. Therefore, an integration of the works of structural equation modeling and item responses theory is required. Considering the extension discussed in both methods as crucial, Raju, N. S., Laffitte, L. J., and Byrne, B. M. (2002)

provide both a theoretical as an exemplified case of the similarities between the confirmatory factor analysis, used to assess a structural equation model, and an item response theory model. Also, the paper discusses the differences between the methods. Based on the similar notions,Glöckner-Rist, A., and Hoijtink, H. (2003) produced a more well-integrated model combining the best assets of both models.

The results of the eventual integrated models must be assessed, but more importantly compared to already known literature. Then, since the models rely upon the controversial p-values, ethical procedures in statistical reporting must be considered. Schumm, W.R.,Pratt, K.K. and Hartenstein, J.L. (2013) discuss the ethics concerning statistical significance. Since working with confidence intervals is possible for the statistics conducted, a shift to the more modern statistical methods of reporting prevents the possibility of controversy.

2.3.6 Discussion and conclusion

The methods of structural equation models have explicitly been listed in order to give a detailed account for the most relevant aspects. A possible pitfall is the mathematical foundation of the methods of structural equation model, as described in one of the previous sections. By conducting the specialized papers described above, the risk of resorting to an ill-defined mathematical framework is minimized. For the IRT-models, the choice of models heavily depends upon the context. Based on the studies already conducted, the partial credit model and the graded response model seem most suitable for this specific study.

Including the relevant theory for integrating both models to answer the main research question concerning the overall, or total, perception of the music displayed is done by the estimation of factor scores. Caution is advised for using the method, although a full exclusion is not necessary. Then, by conducting statistical inferences of the factor scores, the final research question can be answered and simultaneously transformed into relevant information for the theatre.

3 Outlining the Methods of Structural Equation Modeling

A bounteous amount of questions has been rewarded to this thesis. Although cannot be regarded as *'Big data'*, constructing appropriate statistical models for over 150 variables is sinecure. Nonetheless, the data can be secluded across the two discussed dimensions (categorization over clusters and the categorization over time). By epitomization, the questionnaire contains questions related to mood, such as the level of tiredness, happiness and joy. From an informative standpoint, modeling mood in general instead of each specific emotion seems only reasonable. Due to the known structure as exemplified above, the method opted into is **structural equation modeling** (from now on SEM). The method uses a wide variety of mathematical fields and techniques, such as factor analysis, path analysis and measurement models. The most relevant theories within SEM will be examined, employed and illustrated by means of the data.

3.1 Introduction to SEM

Structural equation modeling is a technique that can be traced back to the nineteenth century. The ideas and principles are based on genetic path modeling, a theory introduced by the mathematician Sewall Wright. Within the last fifty to sixty years, the method of SEM has grown in popularity. Sociology and psychometry are the most well-known examples of the usage or the theory, whereas within the field of pure mathematics, the method is not widely applied. Most frequently, the method is used in order to assess a hypothesized structure of the data, while addressing so-called **latent variables**. To illustrate the term informally by an example, consider the quality of a new employee. Since directly measuring such a variable is not feasible, it is measured by means of other observed variables, such as education, pre-existing experience and social capabilities. Then, the unobserved variable *quality* is what is meant by a latent variable. Additionally, the model evaluates the quality of each specific item measurement the *quality*, providing an instrument to continuously improve measurements.

A structural equation model can be broken up into two components (Bollen, 1989):

- the structural model relating the variables and identifying the dependencies of variables; and
- the measurement model identifying the interdependencies with the latent variables and the observed variables

Identifying the theory of factors and exploring covariance structures assembles an understanding of the data. An informed reader may argue that SEM builds on the methods of confirmatory factor analysis. However, for the sake of constructing a broad and well-balanced view, complementary theories such as exploratory factor analysis will be highlighted. In the next section, the notion of a latent variable will be formalized

3.2 Latent variable models

The formulation of a latent variable has already been discussed in chapter two. However, this section tries to rigorously define the latent variable by means of a **latent variable model.**

Definition. [Latent variable model] A *Latent variable model* is a distribution f over the variables X, Y where X are the observed variables and Y are the unobserved, or *latent*, variables.

Example. Although it might seem outside the scope of this thesis, latent variable models are frequently applied in machine learning algorithms, where the *Gaussian mixture model* is the most popular of them. Before discussing the extensiveness of the mathematics, understanding their importance and applications for this thesis must be discussed. More specifically, for the example below, the variables *Are you in a Hurry?* and *Are you excited?* are used.

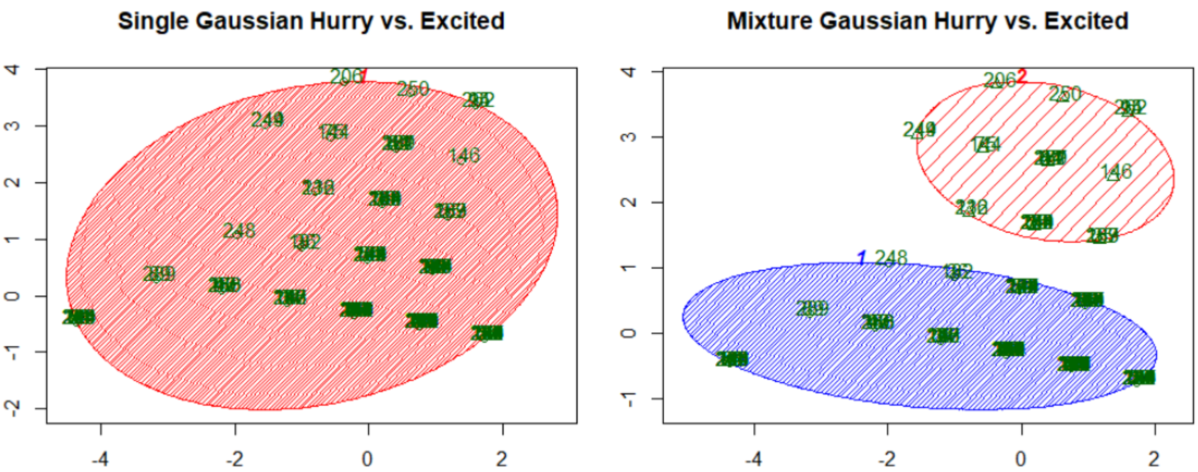

Figure 1: Single Gaussian vs. Mixture of Gaussians ($k = 2$)

For the specific case displayed by the figure above, the cluster acts as an latent variable making statements of performing regression on the data more suitable.

Intermezzo: the numerical strategies involved.

Understanding the structure of data, performing data reduction and conducting item response theory all depend upon the *learning* of a latent variable model. Performing inferences and estimation frequently involves the expressions of the log-likelihood. However, in contrast to maximizing the regular $\log f(x;\theta)$, the log-likelihood of the marginal distribution must be maximized,

$$\max_{x\in D} \log f(D) = \sum_{x\in D} \log\left(\sum_{z} f(x|y)f(y)\right),$$

with D being the data. Due to the inclusion of a prior, two difficulties are posed by the structure of the maximization problem K Ganchev, B Taskar, J Gama - Advances in neural information, 2008) :

1. A summation within the log-operator forbids a well-defined decomposition of $f(x)$ into the summation of log-factors.
2. the right hand side of the expression indicates that the model is a mixture of distributions $f(x|y)$ with corresponding weight $f(y)$. A single exponential distribution must have a likelihood having the concave property (e.g. Siva Balakrishnan, 2019). However, by the introduction of weights, the concave property does no longer apply

Fortunately, the problem at hand has been widely recognized and solved by means of the *Expectation-Maximization algorithm.* Addressing the points made above respectively, the principals can be captured by

1. If y is fully observed, then the optimization of the log-likelihood using $f(x, y)$ is possible
2. Considering the weights $f(y)$ as known, the posterior $f(y|x)$ is computable.

The procedure is conducted as follows. Let X be the set observed variables, Y be the latent variables and let θ be the vector of unknown parameters

Algorithm 1 EM

```
procedure EXPECTATION-MAXIMIZATION(EM)
    L(θ; X, Y) = f(X, Y|θ)
    while Tolerance not attained do
        Ê[θ|θ^(t)) := E_{Y~f(y|x;θ^(t))} log f(x, y; θ)
        θ^(t+1) = arg max_θ Ê[θ|θ^(t)]
    end while
    return θ^(t)
end procedure
```

The algorithm is not only used for Gaussian mixture models and clustering, but is used in the estimation process both used in most notably item response theory. Understanding numerical problems and numerical solutions provides more insight in modeling difficult mathematical concepts, such as a latent variable.

3.3 Factor Analysis

Multivariate statistical models pose a valid method for data reduction, presenting a method to examine the first aim of this study. Factor analysis tries to coalesce various variability-related variables into factors. More formally, factor analysis can be captured into the following definition.

Definition. *Factor Analysis* is the statistical model in which we describe the variability among observed and correlated variables in terms of a lesser number of latent variables, which we shall call *Factors*. More concretely, factor analysis assumes that for a random vector X with $\dim[X] = d$, X consists of correlated variables which can be captured in the factors Y with dimension d_F such that $d_F \leq d$

The observed variables are modelled by representing the variables in terms of a linear combination of the other unobserved variables, plus some noise ε_i. Note that $\varepsilon \sim N(0, \sigma^2)$. Define X_i as the observed variables, Y_i as the factors, where $\#F < \#X$, and let λ_i be the coefficients of the factors, usually captured by the self-explanatory term **Factor loadings**. Then,

$$\begin{aligned}
X_1 &= \mu_1 + \lambda_{11}Y_1 + \lambda_{12}Y_2 + \cdots + \lambda_{1k}Y_k + \varepsilon_1 \\
X_2 &= \mu_2 + \lambda_{21}Y_1 + \lambda_{22}Y_2 + \cdots + \lambda_{2k}Y_k + \varepsilon_2 \\
&\vdots \\
X_n &= \mu_n + \lambda_{n1}Y_1 + \lambda_{n2}Y_2 + \cdots + \lambda_{nk}Y_k + \varepsilon_n
\end{aligned}$$

After centering,

$$\begin{aligned}
X_1 - \mu_1 &= \lambda_{11}Y_1 + \lambda_{12}Y_2 + \cdots + \lambda_{1n}Y_k + \varepsilon_1 \\
X_2 - \mu_2 &= \lambda_{21}Y_1 + \lambda_{22}Y_2 + \cdots + \lambda_{2n}Y_k + \varepsilon_2 \\
&\vdots \\
X_n - \mu_n &= \lambda_{n1}Y_1 + \lambda_{n2}Y_2 + \cdots + \lambda_{nn}Y_k + \varepsilon_n
\end{aligned}$$

or in matrix form

$$\begin{aligned}
\boldsymbol{X} - \boldsymbol{\mu} &= \boldsymbol{\Lambda}\boldsymbol{Y} + \boldsymbol{\varepsilon} \\
\boldsymbol{X}_{\text{cen}} &= \boldsymbol{\Lambda}\boldsymbol{Y} + \boldsymbol{\varepsilon}
\end{aligned}$$

With definitions and dimensions

- $\boldsymbol{X}_{\text{cen}}$: $(d \times 1)$ being the centered random variable;
- $\boldsymbol{\Lambda}$: $(d \times d_F)$ being the factor loadings, with each λ_{ij} is the ith variable of the jth factor. $i = 1, \dots, p$; $j = 1, \dots, F$;
- $\boldsymbol{Y}$: $(d_F \times 1)$ being the latent variables; and
- $\boldsymbol{\epsilon}$: $(d \times 1)$ being the error measurements.

The following assumptions are considered:

- rank$\boldsymbol{\Lambda} = d_F$, id est $\boldsymbol{\Lambda}$ is full rank.
- $Y_i \sim N(0, \boldsymbol{\Phi})$: each Y_i of $\boldsymbol{Y}$ is normally distributed
- $\boldsymbol{\epsilon_i} \sim N(0, \boldsymbol{\Omega})$. Each error is normally distributed. We assume a diagonal Ω with positive elements, or in other words, $\boldsymbol{\Omega} =$diag$(\omega_{11}, \dots, \omega_{dd})$ where each $\omega_{jj} > 0$.

Intermezzo: Why are the error terms normally distributed?

In the definition of a factor model, it was assumed that $\varepsilon \sim N(0, \sigma^2)$. Although this is standard practice, it might be worthwhile to discuss why this is assumed. First of all, as is usually the case, the convenience of modeling with the normal distribution makes a great case of assuming normality. However, mathematical convenience may not be the only reason to use this assumption. By the *central limit theorem*, it can be derived that all the individual errors will tend to a normal distribution with zero mean (e.g. Sang Gyu Kwak Jong Hae Kim, 2017).

Element-wise, each x_i is of the form $x_i = \sum_k \lambda_{ik} y_k + \varepsilon_k$. Considering an example with two Factors Y_1 and Y_2, the assumptions of independence and relationship described above can be visualized by

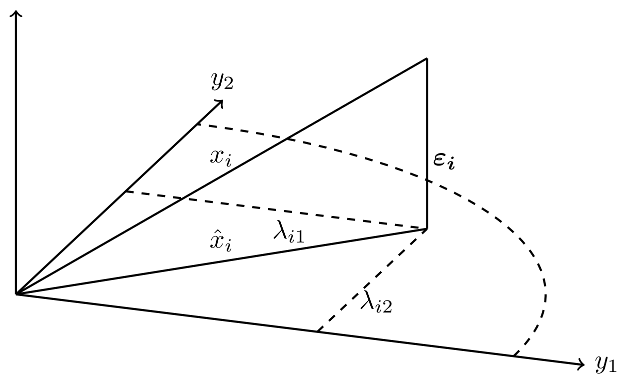

Figure 2: Visual representation of the model

3.3.1 Dimensionality

The data for this study is equipped with pre-defined factors. In other words, it is already known which questions can be clustered together. However, this might not always be the case. Then, the theory of **exploratory factor analysis** seeks the possible factors explaining the variances of its underlying items. Since the factors are already considered known, the method might not seem to be immediately applicable. This assertion is not necessarily true as the mathematical technique also incorporates a method of determining the number of possible factors. The power of eigenvalues are once more demonstrated within this theory, as most methods revolve around criteria based on these eigenvalues.

Excluding the presupposed factors modeling the data, employing an exploratory factor analysis model contributes to an insight of the data. While Kaiser's criterion, in which the number of factors are equal to the number of eigenvalues larger than 1, poses a suitable method, another method is used more frequently (J. Hayton, D Allen, V Scarpello, 2010). Parallel Analysis compares the eigenvalues of the covariance matrix with the eigenvalues of the covariance of a Monte-Carlo process-generated sample of the data. More precisely, the process of a parallel analysis can be captured into the following steps: (J. Hayton, D Allen, V Scarpello, 2010)

1. First, choose a random subset of the data.
2. Then, compute the eigenvalues of the correlation matrix of the sample data. Repeat this process at least fifty times (rule of thumb, J. Hayton, D Allen, V Scarpello, 2010).
3. Thirdly, compute the mean, the 0.95 percentiles of the computed eigenvalues.
4. Next, compare the actual data with the generated samples.
5. Finally, only retain the factors having eigenvalues greater than those of the sampled data.

Computing both metrics for the data, it can be found that

Method	Number of factors
Kaiser's Criterion	32
Parallel Analysis	9

Table 1: Number of factors for the data

In chapter 2, the hypothesized structure has been given. It can be obtained that eleven factors are assumed to model the data, in which three factors are decomposed in time since they contain repeated measures. Therefore, the first informal exploratory analysis is in line with the hypothesized number of factors. Observe that comparing exploratory factor analysis with the researchers beliefs is not a form of confirmatory factor analysis. Nevertheless, it gives a process scrutinizing the proposed number of factors. The method of exploratory factor analysis can also be visualized by means of a simple plot. First, consider the same factor analysis method as described in the previous section

$$\boldsymbol{X} = \boldsymbol{\Lambda Y} + \boldsymbol{\varepsilon}$$

Then, if $\boldsymbol{T}$ designates the rotation matrix defined by

$$\boldsymbol{T} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$

the rotated factor loadings matrix is denoted by

$$\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda T}$$

such that

$$\boldsymbol{X} = \boldsymbol{\Lambda}^* \boldsymbol{Y} + \boldsymbol{\varepsilon}$$

Demonstrating the application of such a rotation becomes clear whenever a visualisation is made as below.

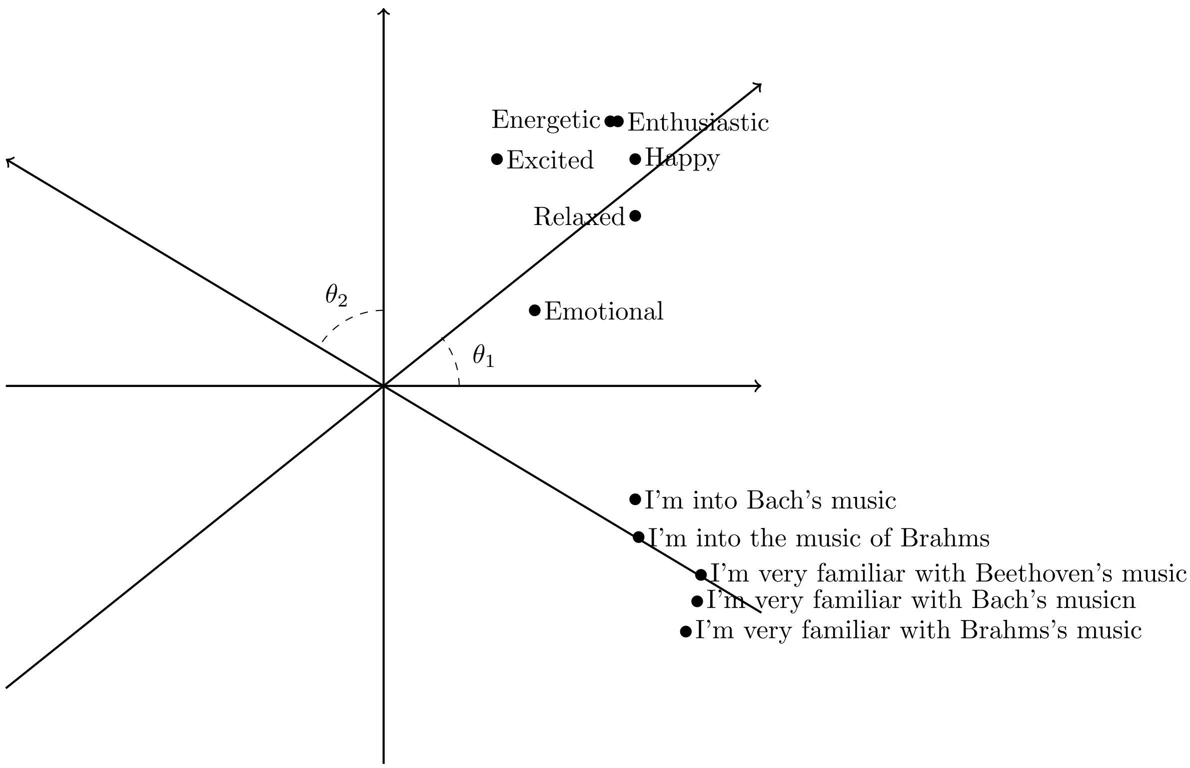

Figure 3: Exemplifying factor analysis with rotation for the actual data

Note that this plot is based on the actual data. Visually, an evident pattern is recognizable. Since mathematical techniques have more difficulty with making use of this pattern due to the standard orientation of the axis, the insertion of a rotation matrix models the pattern. Since the number of factors from the analysis are relatively close to the hypothesized ones (nine vs. eleven), the next step is assessing the items which load onto each of the eleven factors.

3.4 Confirmatory Factor Analysis and Covariance

In contrast to exploratory factor analysis, confirmatory analysis assumes a known structures which needs to be tested. The model is of the form

$$\boldsymbol{X} = \boldsymbol{\Lambda Y} + \boldsymbol{\varepsilon}$$

where Y are the observed variables, Λ is the so-called factor loadings matrix for the unobserved, or latent, variables ξ with ε as the disturbance term. The theory of a confirmatory factor analysis model will be discussed in the next chapter. However, the representation displayed above will be useful for introducing the notion of covariance and the possibility to model by means of covariance. Identification and estimation of structural models are based on the covariance matrix. This section tries to both formally discuss the covariance matrix, after which a demonstration of covariance is given. To comply with mathematical standards, the formal definition of the covariance operator is given by

Definition. On a Hilbert space H with inner product $\langle \cdot, \cdot \rangle$ and the probability measure $\mathbb{P}$, the *Covariance operator* $H \times H \to \mathbb{R}$ is given by

$$\mathrm{cov}(x, y) = \int_H \langle x, z \rangle \langle y, z \rangle \, \mathrm{d}\mathbb{P}(z)$$

More concretely, the covariance matrix of a random vector $\boldsymbol{X} = [x_1, \ldots, x_d]^T$, each with well-defined variance and expectation, is defined as

$$\boldsymbol{\Sigma} = \mathrm{cov}(X_i, X_j) = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])]$$

For the confirmatory factor analysis model described above, a simple computation by using the definition above yields the covariance matrix (Note that $\mathbb{E}[\boldsymbol{X}_{\text{cen}}] = 0$)

$$\begin{aligned} \boldsymbol{\Sigma} = \mathbb{E}[(\boldsymbol{X}_{\text{cen}})(\boldsymbol{X}_{\text{cen}})^T] &= \mathbb{E}[(\boldsymbol{\Lambda Y_i + \epsilon_i})(\boldsymbol{\Lambda Y_i + \epsilon_i})^T] && (1) \\ &= \boldsymbol{\Lambda}\mathbb{E}(Y_i Y_i^T)\boldsymbol{\Lambda^T} + 2\boldsymbol{\Lambda}\mathbb{E}[Y_i \varepsilon_i^T] + \mathbb{E}[\varepsilon_i \varepsilon_i^T]. && (2) \end{aligned}$$

As $Y_i \perp\!\!\!\perp \varepsilon_i$, it follows $\boldsymbol{\Lambda}\mathbb{E}[Y_i \epsilon_i^T] = 0$. Also, let $\boldsymbol{\Phi} = \mathbb{E}(Y_i Y_i^T)$ and $\Psi = \mathbb{E}[\varepsilon_i \varepsilon_i^T]$. Then,

$$= \boldsymbol{\Lambda}\Phi\boldsymbol{\Lambda} + \Psi \qquad (3)$$

Corollary. The covariance matrix of $\boldsymbol{X}$ is a positive semi-definite matrix for all $a \in \mathbb{R}^d$.

Proof: A matrix is positive definite whenever, for $a \in \mathbb{R}^d$ and square $d \times d$ matrix X if the scalar $a^T X a \geq 0$. By following this definition, the computation of the scalar $a^T \boldsymbol{X} a$ yields

$$\begin{aligned} a^T \boldsymbol{\Sigma} a &= a^T \mathbb{E}[(\boldsymbol{X}_{\text{cen}})(\boldsymbol{X}_{\text{cen}})^T] a \\ &= \mathbb{E}[a^T (\boldsymbol{X}_{\text{cen}})(\boldsymbol{X}_{\text{cen}})^T a] \\ &= \mathbb{E}[(a^T \boldsymbol{X}_{\text{cen}})^2] \\ &\geq 0. \end{aligned}$$

3.4.1 Understanding Covariance and Eigendecomposition

Formal definitions aside, understanding the importance and aim of modeling by covariance yields an intuitive view of the subject. Considering the variance as a the 'spread' of the data, the covariance of two random variables is less straightforward to capture in one informal word of sentence. To illustrate the notion of covariance, consider two dimensional data. In this case, the variables *"Knowledge of the music of Bach"* and *"Knowledge of the music of Brahms"* are taken into account. Intuitively, expecting correlation between the two variables seems only reasonable. A clear diagonal orientation is visible. Calculating the representative covariance matrix yields

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{xx}^2 & \sigma_{xy}^2 \\ \sigma_{yx}^2 & \sigma_{yy}^2 \end{bmatrix} = \begin{bmatrix} 1.65 & 1.14 \\ 1.14 & 1.52 \end{bmatrix} \qquad (4)$$

which is in line with the plot in figure 4. The diagonal elements are equal to the familiar variances of the x and y components. The non-diagonal covariance elements of x, y are nonzero, and hence a positive correlation is detected. Informally, the covariances of x and y can be regarded as the 'orientation' of the data. Following the latter notion, identifying the direction and magnitude of the orientation

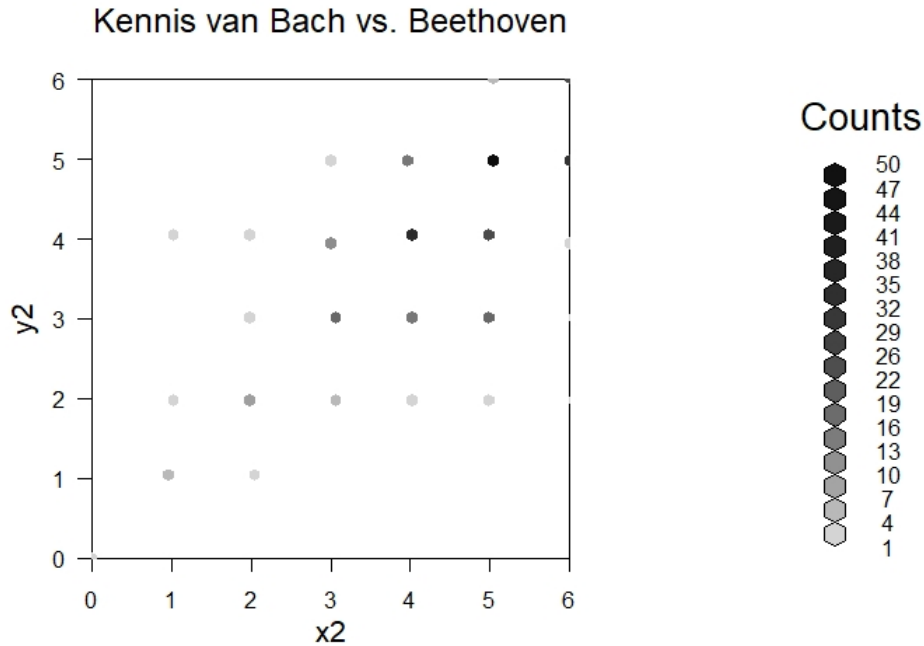

Figure 4: Scatterplot of 2D data

produces relevant information about the data. Define Z as the exemplified two dimensional data, and let $\boldsymbol{v}$ be the direction of the orientation of the data. Then, the covariance has the quadratic form

$$\text{cov}(Z) = \boldsymbol{v}^T \boldsymbol{\Sigma} \boldsymbol{v}$$

The direction and hence $\boldsymbol{v}$ is determined by pointing into the direction of the largest variance. Such a problem can be mathematical stated by finding $\max \boldsymbol{v^t \Sigma v}$. In the topic of eigenvalues of a real symmetric matrix, the **Rayleigh quotient** is commonly employed for this specific maximisation problem

Definition. The *Rayleigh quotient* for a some real symmetric matrix M and some $x \geq 0$ is defined as

$$R(M, x) = \frac{x^T M x}{\|x\|^2}.$$

Theorem. Let M be a real squared symmetric matrix. Then, the largest eigenvalue $\lambda_{\max}$ is obtained by

$$\lambda_{\max} = R(M; x)$$

Proof in the appendix.

Corollary. $\boldsymbol{\Sigma}$ has orthogonal eigenvectors.

Proof: Let $v = \{v_1, \ldots v_n\}$ be the set of eigenvectors, and let $\lambda = \{\lambda_1, \ldots, \lambda_n\}$ be the set of eigenvalues. Then,

$$\begin{aligned}
\boldsymbol{\Sigma} v_i = \lambda_i v_i &\implies v_k^T \boldsymbol{\Sigma}_i = v_j^T \lambda_i v_i \\
&\implies \lambda_j v_j^T v_i = \lambda_i^T v_j v_i \\
&\implies (\lambda_j - \lambda_i) v_j^T v_i = 0 \\
&\implies v_j^T v_i = 0
\end{aligned}$$

In short, the eigenvector associated with the largest eigenvector has direction of the largest variance. For the case at hand, computing the eigenvalues and vectors yields

$$\lambda_i = \{2.79, 0.45\} \quad \& \quad \boldsymbol{e}_i = \left\{ \begin{bmatrix} -0.72 \\ -0.68 \end{bmatrix}, \begin{bmatrix} 0.68 \\ -0.72 \end{bmatrix} \right\}, \quad \text{for } i = 1, 2$$

adjusting these vectors with the representative x and y mean, plotting the eigenvectors results in the plot below.

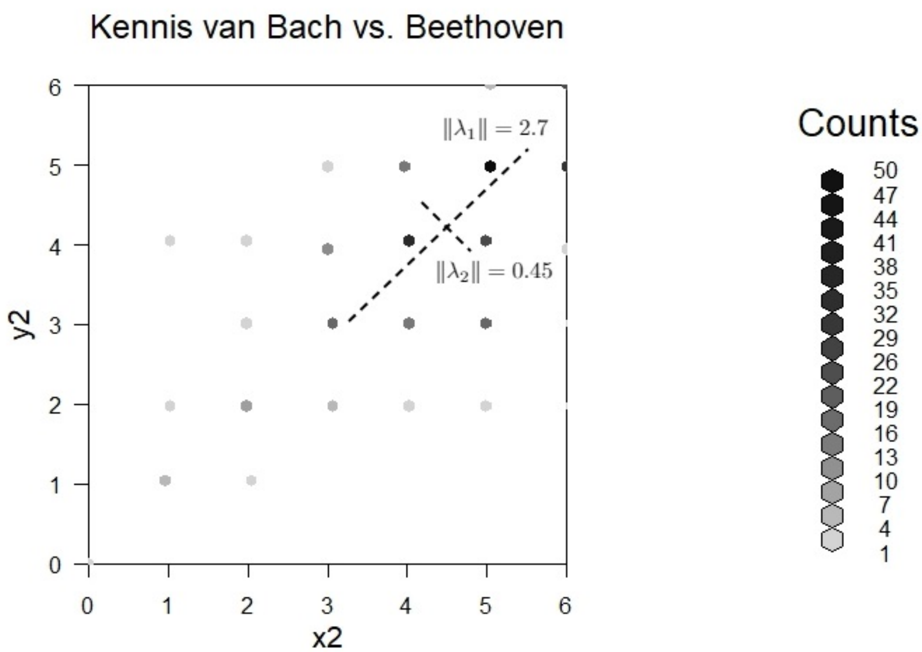

Figure 5: Eigendecomposition of the covariance matrix

Whereas the variance of the two dimensional data describes the variance of the x and y components along the x and y axis, the eigenvalues represent the magnitude of the variance adjusted to the direction having the largest variance. For having a visual counterpart, two dimensional data is best applicable. However, the methods used can easily be extended to having d dimensions.

3.4.2 Estimation

Estimation is a central part of statistics. Point estimation, hypothesis testing and Bayesian inference all depend upon estimation. For the covariance matrices, the method of estimation usually involves **Maximum Likelihood Estimation**. As $\varepsilon \sim N(O, \boldsymbol{\Sigma})$, the starting point revolves around the multivariate normal distribution

$$f(\boldsymbol{z}) = \prod_{i=1}^{p} f(z_i) = \left(\frac{1}{\sqrt{2\pi}} \right)^d \det[\boldsymbol{\Sigma}] \exp\left[-\frac{1}{2}(z-\mu)^T \boldsymbol{\Sigma}^{-1}(z-\mu) \right]. \tag{5}$$

Computing the loglikelihood yields

$$
\begin{aligned}
l(\mu, \mathbf{\Sigma}; x_1, \ldots, x_n) &= \log\left\{\prod_{i=1}^{n} f(\boldsymbol{z})\right\} \\
&= \log\left\{\prod_{i=1}^{n} \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(x_i-\mu)^T\mathbf{\Sigma}^{-1}(x_i-\mu)\right]\right\} \\
&= \sum_{i=1}^{n}\left\{\underbrace{\frac{d}{2}\log(2\pi)}_{\text{constant}} - \frac{1}{2}\log|\mathbf{\Sigma}| - \frac{1}{2}\Big((x_i-\mu)^T\mathbf{\Sigma}^{-1}(x_i-\mu)\Big)\right. \\
&= C - \frac{n}{2}\log|\mathbf{\Sigma} - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T\mathbf{\Sigma^{-1}}(x_i-\mu) \\
&= C - \frac{1}{2}\Big(n\log|\mathbf{\Sigma}| + \operatorname{trace}[S_\mu\mathbf{\Sigma}^{-1}]\Big)
\end{aligned}
$$

where

$$S_\mu = \sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T$$

Handling a likelihood with matrices instead of real numbers is somewhat more complex and requires the following identities

1. $\operatorname{trace} AB = \operatorname{trace} BA$.
2. $x^TAx = \operatorname{trace} x^TAx = \operatorname{trace} x^TxA$
3. $\dfrac{\partial}{\partial A}\operatorname{trace} AB = B^T$
4. $\dfrac{\partial}{\partial A}\log|A| = A^{-T}$

Then

$$\frac{\partial}{\partial A}x^TAx = \frac{\partial}{\partial A}x^TxA = [xx^T]^T = xx^T$$

Implying

$$l(\mu, \mathbf{\Sigma}; x_1, \ldots, x_n) = C + \frac{1}{2}\Big(n\log|\mathbf{\Sigma}^{-1}| + \operatorname{trace}[S_\mu\mathbf{\Sigma}^{-1}]\Big)$$

Computing the derivative with respect to $\mathbf{\Sigma}^{-1}$

$$\frac{\partial l}{\partial\Sigma^{-1}} = 0 + \frac{1}{2}\Big(n\mathbf{\Sigma} + S_\mu^T\Big)$$

Note that S_μ is symmetric. Equating to zero yields

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}S_\mu$$

3.5 The two components of SEM

The framework of Factor Analysis and Covariance is necessary to fully grasp structural equation modeling. Model assessment aside, SEM can be broken down into *measurement models*, modeling latent variables, and *path analysis*, modeling causality between variables.

Definition. An *endogenous* variable, usually denoted by $\boldsymbol{\eta}$, is a variable determined by the relationship with other variables. It's opposite is a *exogenous* variable, being independent from other variables. In the models, it will be denoted by $\boldsymbol{\xi}$

3.5.1 Path Analysis

Path Analysis is a method used within SEM, in which causality is (usually also graphically) modeled. A recursive model of the latent variables listed in the data is

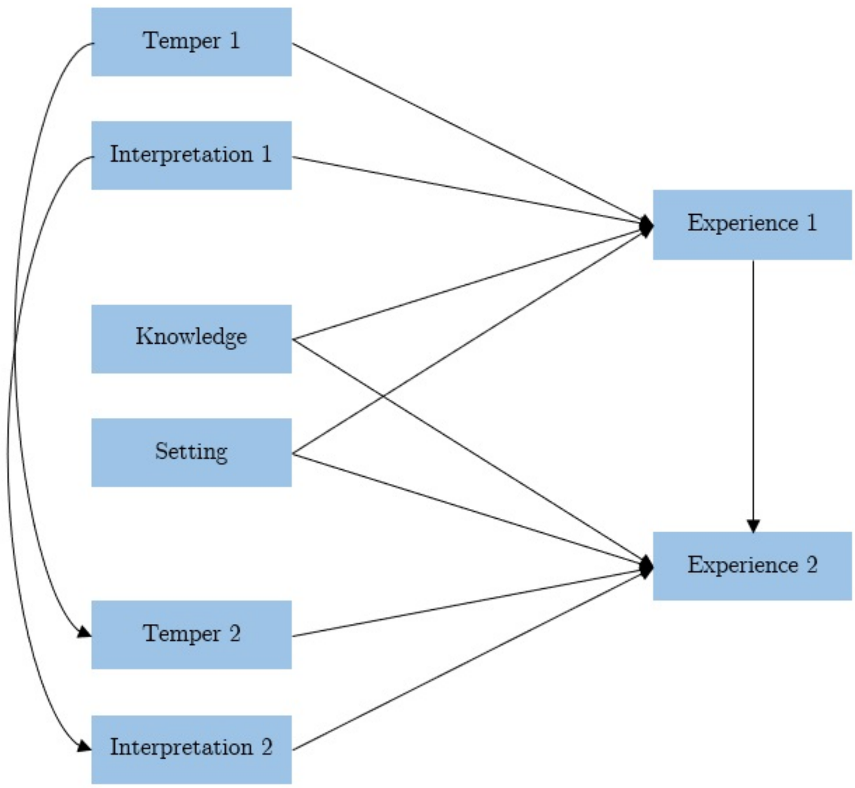

Figure 6: A path model for a subset of the data

The goal is to formalize the model above. First, define

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ \eta_6 \end{bmatrix} = \begin{bmatrix} \text{Experience 1} \\ \text{Experience 2} \\ \text{mood 1} \\ \text{Interpretation 1} \\ \text{mood 2} \\ \text{Interpretation 2} \end{bmatrix}, \quad \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} \text{Knowledge} \\ \text{Setting} \end{bmatrix}.$$

The structural equations writes

$$\eta_1 = \beta_3\eta_3 + \beta_4\eta_4 + \gamma_1\xi_1 + \gamma_2\xi_2 + \varepsilon_1$$
$$\eta_2 = \beta_1\eta_1 + \beta_5\eta_5 + \beta_6\eta_6 + \gamma_3\xi_1 + \gamma_4\xi_2 + \varepsilon_2$$

Or in matrix form

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\varepsilon} \tag{6}$$

The recursive structure is only logical to implement, as the experience at time 1 intuitively correlates with the experience at time 2. The same principle holds for other variables measured at different instances of time. Nonetheless, the asserted structure cannot perfectly fit the model. Therefore, an error term $\boldsymbol{\varepsilon}$ is incorporated. Similar to the factor analytic model, strict assumptions are imposed on the model assuring valid models.

- For the latent variables, $\boldsymbol{\eta} = \boldsymbol{\eta}_{\text{cen}}$, meaning that y is centered around its mean, and thus $\mathbb{E}[\boldsymbol{\eta}] = 0$. The same holds for $\boldsymbol{\xi}$, thus $\mathbb{E}[\boldsymbol{\xi}] = 0$
- The error terms are normally distributed: $\boldsymbol{\varepsilon} \sim N(O, \boldsymbol{\Psi})$
- The model is recursive in the sense that an element of η may dependent upon **another** element or elements of η. However, the specific element of η may not depend on itself. Therefore, the diagonal elements of $\boldsymbol{B}$ are all set to zero.

As the structural model tries to elucidate the variable y in terms of exogenous and endogenous variables, the variance of the exogenous variable $\boldsymbol{\xi}$ measures the explained variable by means of those exogenous variables related to the endogenous variable $\boldsymbol{\eta}$. The error term $\boldsymbol{\varepsilon}$ has covariance matrix Ψ, displaying the unexplained variance of the model.

3.5.2 Measurement models

The second central topic within structural equation modeling comprises the measurement model, identifying the measurement of the latent variables. A tricky but necessary assumptions, which heavily relies on the assessment of the researcher, is the proper correlation of latent variables and its respective or assigned observed variables. The general model for a structural equation models writes

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\varepsilon}$$

Both $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ contain latent variables which are required to be measured. For this, the linear relationships

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{\Lambda_x}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \\ \boldsymbol{y} &= \boldsymbol{\Lambda_y}\boldsymbol{\xi} + \boldsymbol{\delta} \end{aligned}$$

are employed. To exemplify the notation, consider a very simple model with the latent variables *mood* at time 1 and 2 (η_3 and η_5 respectively from the previous subsection). Let

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} \text{Energetic } t_1 \\ \text{Enthusiasm } t_1 \\ \text{Happiness } t_1 \\ \text{Relaxed } t_1 \\ \text{Emotional } t_1 \\ \text{Excited } t_1 \end{bmatrix}, \qquad \begin{bmatrix} x_7 \\ x_8 \\ x_9 \\ x_{10} \\ x_{11} \\ x_{12} \end{bmatrix} = \begin{bmatrix} \text{Energetic } t_2 \\ \text{Enthusiasm } t_2 \\ \text{Happiness } t_2 \\ \text{Relaxed } t_2 \\ \text{Emotional } t_2 \\ \text{Excited } t_2 \end{bmatrix},$$

where $x_1, \ldots, x_6$ are the items measuring latent eta_3 and $x_7, \ldots, x_{12}$ are the items measuring η_5. Then, component-wise,

$$\begin{aligned} x_1 &= \lambda_1 x_3 + \varepsilon_1 & x_7 &= \lambda_7 x_5 + \varepsilon_7 \\ x_2 &= \lambda_2 x_3 + \varepsilon_2 & x_8 &= \lambda_8 x_5 + \varepsilon_8 \\ x_3 &= \lambda_3 x_3 + \varepsilon_3 & x_9 &= \lambda_9 x_5 + \varepsilon_9 \\ x_4 &= \lambda_4 x_3 + \varepsilon_4 & x_{10} &= \lambda_{10} x_5 + \varepsilon_{10} \\ x_5 &= \lambda_5 x_3 + \varepsilon_5 & x_{11} &= \lambda_{11} x_5 + \varepsilon_{11} \\ x_6 &= \lambda_6 x_3 + \varepsilon_6 & x_{12} &= \lambda_{12} x_5 + \varepsilon_{12}. \end{aligned}$$

or, in matrix form,

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_{12} \end{bmatrix} \quad \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ \vdots & \vdots \\ \lambda_6 & 0 \\ 0 & \lambda_7 \\ \vdots & \vdots \\ 0 & \lambda_{12} \end{bmatrix}, \quad \boldsymbol{\eta} = \begin{bmatrix} \eta_3 \\ \eta_5 \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{12} \end{bmatrix}.$$

Note, it is also possible to restrict the factor loadings λ such that $\lambda_1 = \lambda_7, \ldots, \lambda_6 = \lambda_{12}$. Then, the factor have a similar loading onto the latent variable. This restriction is reasonable to impose whenever it is believed that item have to load similarly onto a latent variable. For this study, this is also the case. Although *mood* at time 1 and 2 might be different valued, their respective factor loadings should be equal since they measure the exact same items.

3.5.3 Estimation of parameters of the structural equation model

A few sections preceeding this section, the ML-estimate for the covariance matrix was discussed. Now, as the two components of the structural model have been given, the next step is to seek a suitable estimator for the model. Multiple estimators can be formed for the model, but most show lackluster properties. Although the *unweighted least squares* is an intuitively pleasing estimator, the method shows disadvantages, such as lacking the property of being scale invariant. This is no problem if the model at hand admits a uniform scale, but this is not the case. Therefore, the F_{ULS} method cannot be considered for this project. For solving this problem, the inclusion of a weight matrix is implemented in the *weighted least squares* method

$$F_{ULS} = \frac{1}{2} \operatorname{trace} (\boldsymbol{W}^{-1}(\boldsymbol{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta})))^2$$

where the weights are derived from the residuals. This method takes into account possible heteroscedasticity or auto-correlation between the residuals, which is applicable to the model. However, the most considered method is the maximum likelihood estimate, showing maximum efficiency and consistency and it is given by

$$F_{ML} = \log |\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \operatorname{trace} \boldsymbol{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \log |\boldsymbol{S}| - q \tag{7}$$

where q denotes the rank of $\boldsymbol{\Sigma}$.

4 The structural model

The methods of a structural equation model has had a subtle introduction in the previous sections. Such a model is, however, not merely a method of presenting a plain structure to some factor analysis-esque equations, but provides a method of assessing a hypothesized structure of the data, identifies alternative structures and consists of an estimation procedure thus providing a numerical interpretation to the unobserved latent variables. A complete structural equation model consists of a **latent variable model** and a **measurement model**, as discussed in chapter 3. This section integrate the two components into a absolute model, after which its identification will be established. The first half of the model is constructed by the latent variable model

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}$$

where the notion is in line with Bollen (1989), with the following definitions

- $\boldsymbol{\eta}$ is the vector of latent endogenous variables with dimension $d \times 1$.
- $\boldsymbol{\xi}$ are the latent exogenous variables with $n \times 1$
- $\boldsymbol{B}$ includes the factors loadings for the interdependence of the endogenous latent variables
- $\boldsymbol{\Gamma}$ includes the factor loadings for the interdependence of the exogenous latent variables.

The next component comprises the system of measurement models:

$$\boldsymbol{Y} = \boldsymbol{\Lambda_y}\boldsymbol{\eta} + \boldsymbol{\epsilon}$$
$$\boldsymbol{X} = \boldsymbol{\Lambda_x}\boldsymbol{\xi} + \boldsymbol{\delta}$$

with definitions

- y with dimensions $p \times 1$ and x with dimension $q \times 1$ are the observed variables
- Λ_y with dimensions $p \times m$ and Λ_x with dimensions $q \times n$ are the coefficient matrices.
- ϵ, δ are the matrices concerning the error terms.

Note the similarity with a confirmatory factor analysis model. By a simple computation, Bollen (1989) showed that the *implied* covariance matrix, based on the structural model, is given by

$$\boldsymbol{\Sigma}(\boldsymbol{\Omega}) = \begin{bmatrix} (\boldsymbol{I}-\boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}^T + \boldsymbol{\Psi})((\boldsymbol{I}-\boldsymbol{B})^{-1})^T & (\boldsymbol{I}-\boldsymbol{B})^{-1} \\ \boldsymbol{\Phi}\boldsymbol{\Gamma}^T((\boldsymbol{I}-\boldsymbol{B})^{-1})^T & \boldsymbol{\Phi} \end{bmatrix}$$

Constructing statistical inferences of a structural model requires a proper method of estimation for the covariance matrix. The process pointing out whether estimation is well defined is called **identification**.

4.0.1 Identification

Almost every equation and system has restrictions to be considered well-defined. Within the methods of SEM, this will be referred to as **identifiability**.

Definition. A parameter vector Ω is not globally identified if $\boldsymbol{\Sigma}(\boldsymbol{\Omega}_1) \neq \boldsymbol{\Sigma}(\boldsymbol{\Omega}_2) \implies \boldsymbol{\Omega}_1 \neq \boldsymbol{\Omega}_2$.

Observe the argumentative contra-positivity in the definition. Bollen (1988) provides a diverse set of identification metrics. Nonetheless, they all require a diagonal $\boldsymbol{\Psi}$, which does not resemble the data in which items from different latent variables are correlated (the items of *mood* of time 1 and 2 are component-wise correlated). To account for this problem, several checks are possible for checking mathematical and numerical criterions.

t-rule A model based on matrices is not solvable if the number of estimated parameters exceeds the number of unique parameters. More specifically, the number of unique elements of a matrix is equal to $q \cdot (q+1)$. Note that $\boldsymbol{\Sigma}$ is symmetric, and hence the unique elements reduce to $\frac{1}{2}q(q+1)$. Defining t to be the unknowns, it is required that $t \leq \frac{1}{2}q(q+1)$. Due the contra-positive nature of the argument, the t-rule gives a check for mis-identification, and is consequently a necessary- but not a sufficient condition for identifiability.

Example. Consider the two factor model for experience at time 1 and 2. By using the decomposition discussed in the previous chapter, the number of estimated parameters

$$\mathbf{\Sigma}(\mathbf{\Omega}) = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}^{\boldsymbol{T}} + \mathbf{\Psi}$$

is sought. By an ordinary substitution

$$\Gamma = \begin{bmatrix} \lambda_1 & 0 \\ \vdots & \vdots \\ \lambda_6 & 0 \\ 0 & \lambda_7 \\ \vdots & \vdots \\ 0 & \lambda_{12} \end{bmatrix}, \quad \Phi = \begin{bmatrix} \phi_{11} & \\ \phi_{12} & \phi_{22} \end{bmatrix} \tag{8}$$

Now, the matrix $\mathbf{\Psi}$ containing the measurement error is the only matrix left to identify. Note that is only reasonable to let δ_1 correlate to δ_7 due to the correlation of x_1 and x_7. Therefore,

$$\Psi = \begin{bmatrix} \mathrm{V}(\delta_1) & & & & & & & & \\ & \mathrm{V}(\delta_2) & & & & & & & \\ & & \mathrm{V}(\delta_3) & & & & & & \\ & & & \mathrm{V}(\delta_4) & & & & & \\ & & & & \mathrm{V}(\delta_5) & & & & \\ & & & & & \mathrm{V}(\delta_6) & & & \\ \mathrm{Cov}(\delta_1, \delta_7) & & & & & & \mathrm{V}(\delta_7) & & \\ & \ddots & & & & & & \ddots & \\ & & \mathrm{Cov}(\delta_6, \delta_{12}) & & & & & & \mathrm{V}(\delta_{12}) \end{bmatrix}$$

Then, by counting the unknowns in all matrices, we have that $1/2(q(q+1)) = 78, t = 33$ resulting in the possibility of identification. Numerical necessity of identification should also be part of this section as all methods used heavily rely on computer algebra.

Definition. Let $f : \mathbb{R} \mapsto \mathbb{R}^n$ be a scalar valued function with well defined second partial derivatives. Then, the *Hessian* matrix $\boldsymbol{H}$ is defined by

$$\boldsymbol{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial^2 x_n} \end{bmatrix}.$$

For a feasible computation, the Hessian of the likelihood also used in the determination of the maximum likelihood estimator is required to be positive definite, posing another check for the identification for the model.

4.1 Mathematical soundness

The application of a structural equation model is easily done by a decent number of mathematical tools (e.g. SPSS, Stata, R). Nevertheless, without a sufficient understanding of the mathematical techniques, assumptions constructing the model are easily violated (West, Stephen G. Finch, John F. Curran, Patrick J, 1995). This section tries to perspicaciously identify each assumption or statistical pitfalls after which a possible mathematical solution will be given and demonstrated. The upcoming subsections list the main problems within the framework and consists of problems with non-normality, specification error and the seemingly unrestricted construction of exogenous variables.

4.1.1 Non-normality

Assuming normal distributed data is a very common practice in statistics, but may sometimes be an ill-considered premise. Questions related to the possible normality of categorical variables, the violation of the assumption and the estimation procedure based on this assumption arise.

Problem. The data may not always be normally distributed.

Consequences. The estimation procedure is based on the data having a normal distribution. Therefore, the estimated parameters might be inappropriate. However, violation of the normality assumption shows no problem for the parameter estimation, but produces issues for the efficiency of the standard errors. Consequently, the standard errors of the current estimation cannot be guaranteed to be the smallest. Also Boomsma (1983) states that the violation of the normality assumption leads to the overestimation of the likelihood ratio chi-square statistic.

Solution. To account for the non-normality, the least squares estimation should be replaced by the weighted least squares method to account for a non-normality in the error terms (or: heteroscedasticity). It should be noted that the study works with discrete data, which can never follow a normal distribution, as a normal distribution is continuous. However, by inspection of histograms, such as the histogram below, it can be obtained that almost every data-point seems approximately normal, with usually a right or left skew. Therefore, the weighted least squares will be used for the estimation procedure.

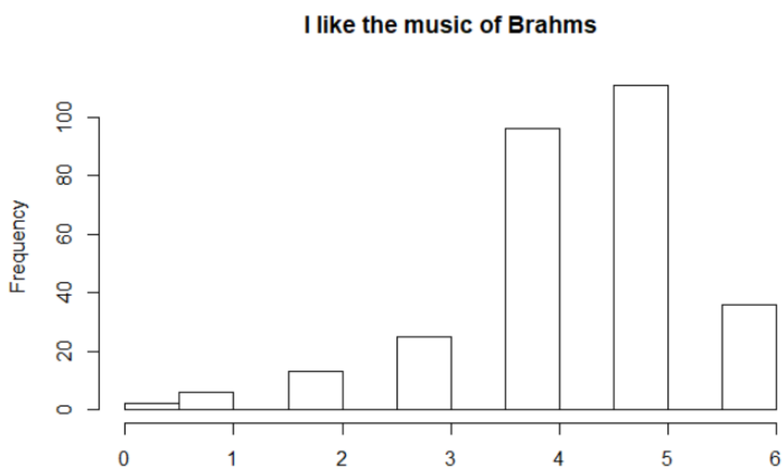

Figure 7: Histogram of *I like the music of Brahms*

4.1.2 Specification error, modification index and the Wald test

A structural equation model assumes a measurement error in the measurement model. Although a fully correct fit is not to be expected, other errors also arise, such as the *specification error*.

Problem. The *specification error* is the omission of other possibly necessary variables in any of the measurements of any of the structural equations.

The question arises what the influence of this error is in the model.

Consequences. The structural equation model omits a non-trivial parameter estimate bias, as discussed in Kaplan (1988). The more important consequence is inherent to its name, as the model is not well specified.

Solution. The solution for this situation requires more insight into the current and most used methods of correction of specification errors. It can be divided into two steps:

1. The application of the chi-square test, testing for amendments of the model and its necessity
2. After it is concluded that the model should be modified, the *Modification Index* is used to identify which restrictions should be changed.

The assessment of the statistical quality of the model is inherent to the error of specification. The most common method of assessment within the method relies on the chi-square statistic, in line with most other statistical methods. The application of the p-value is not without its controversies (e.g. Amrhein, Greenland and McShane, 2019). Additionally, the chi-square test for SEM-based models pose another problem. Namely, the chi-square statistic and hence the p value is fully proportional to the sample size. Illustrating the latter statement, computation of the likelihood ratio statistic yields

$$L_0 = -\frac{n}{2}\{\log|\boldsymbol{\Sigma}(\boldsymbol{\Omega})| + \text{trace}\, \boldsymbol{S}\boldsymbol{\Sigma}^{\boldsymbol{-1}}\boldsymbol{(\Omega)}\}$$

with alternative that the loglikelihood attains its maximum as $\hat{\Sigma} = S$. Then,

$$L_a = -\frac{n}{2}\log|\boldsymbol{S}| + \text{trace}\, \boldsymbol{S}\boldsymbol{S}^{\boldsymbol{-1}} = \frac{n}{2}\log|\boldsymbol{S}| + q.$$

The likelihood ratio test is based on the relationship

$$\begin{aligned} -2\log\frac{L_0}{L_a} &= n\{\log|\boldsymbol{\Sigma}| + \text{trace}\, \boldsymbol{\Sigma}^{-1}\boldsymbol{S} - \log|\boldsymbol{S}| - q\} \\ &= n \cdot F_{ML}. \end{aligned}$$

Note the sensitivity for n. As pointed out by e.g. Jöreskog and Sörbin (1983), the sample size is directly proportional to the test statistic. The resulting p-value will therefore easily be statistically significant. Paradoxically, for a structural model it is essential to have a sufficiently large sample size (e.g. TA Kyriazos - Psychology, 2018). Hence, a different method of assessment should be used. An alternative method of assessment evaluates the quality of the model specification. The specification of a model is based on restrictions of factor loadings. To exemplify such a restriction, consider the possibility to let $\lambda_{11} = \lambda_{52}$ such that the factor loadings are equal, or $\lambda_{12} = 0$ such that the respective item does not load onto the respective latent variable. Then, by Satorra (1987), it is possible to test the difference in models in which a restriction is imposed or relaxed. Consider the statistic

$$\Delta F = nF^* - nF = n(F^* - F)$$

where F^* denotes the restricted model, and F denotes the more general model. Note that if no restrictions are imposed, $F = 0$. Then,

- If the baseline model F^* is correct and the data omits normality, ΔF will be asymptotically chi square distributed with the difference in known elements as the degrees of freedom
- If the baseline model F is incorrect and the data omits normality, ΔF will be asymptotically non-centrally chi square distributed with the difference in known elements as the degrees of freedom. Satorra and Saris (1985) showed that the non-centrality parameter λ is equal to the test statistic itself, ΔF.

This alternative assessment of the model less dependent upon the sample size concerns the placed restriction upon the model and their validness. The method can also be extended such that it provides information about the possible improvements of the model. Consider the score vector

$$\boldsymbol{U}(\boldsymbol{\Omega}) = \frac{\partial \log L(\boldsymbol{\Omega})}{\partial \boldsymbol{\Omega}}$$

which is simply the first derivative of the log-likelihood. Note that the $U = 0$ if $\boldsymbol{\Omega}$ is chosen to be the maximum likelihood estimator F_{ML} by construction. However, the estimation does not concern the resticted model as exemplified above. Consider $\boldsymbol{\Omega}^*$, which is the restricted vector of parameters of Ω. If Ω^* fits the model perfectly, then U will vanish. However, if U does not vanish under the Ω^*, the restrictions do not hold perfectly. Practically, perfect restrictions are utopian thoughts. The level of discrepancy indicates the appropriateness of the restricted model, and can be composed by the *Modification Index Test* (Silvey 1959)

$$MI = U(\boldsymbol{\Omega}^*)^T (I(\hat{\boldsymbol{\Omega}}^*)^{-1}(U(\hat{\boldsymbol{\Omega}}^*))$$

where $MI \sim \chi^2$ with the number of restrictions as the degrees of freedom. This test is more commonly referred to as the Lagrange Multiplier test, but for the context of this thesis the first term is more fitting. However, the MI-test only accounts for a test for a more restrictive and less restrictive model. We wish to evaluate this principal for a fully unrestricted model and the final restricted model. Then, by first defining $R(\boldsymbol{\Omega})$ to be the set of restrictions. Then, by D Kaplan 2009, the Wald statistic for the assessment of the full model vs. the full restricted model is

$$\boldsymbol{W} = (R(\hat{\boldsymbol{\Omega}}_u))^T \left\{ \left[\frac{\partial R(\hat{\boldsymbol{\Omega}}_u)}{\partial \hat{\boldsymbol{\Omega}}_u} \right] (I(\hat{\boldsymbol{\Omega}}_u)) \left[\frac{\partial R(\hat{\boldsymbol{\Omega}}_u)}{\partial \hat{\boldsymbol{\Omega}}} \right] \right\}^{-1} (R(\hat{\boldsymbol{\Omega}}_u))$$

which again is asymptotically chi-square distributed. The lavaan-software has a built in functions for this method and will be used accordingly.

Another theoretical problem: weak exogeneity.

Another possible problem arises in the designation of a exogenous variable. Recall the model for a structural equation model

$$\boldsymbol{\eta} = \boldsymbol{B\eta} + \boldsymbol{\Gamma\xi} + \boldsymbol{\delta}$$

By the construction, the parameter ξ is exogenous. However, a simple designation may not imply full exogeneity. Explaining the problem in the normal general linearized model context seems appropriate. Let Z be denote the matrix of independent and dependent variables concerned. Or in this case, Z can be partitioned into the endogenous and exogenous variables. The partitioning is also possible from the model theoretic standpoint, since the model can be written as

$$\begin{aligned}
(\boldsymbol{I} - \boldsymbol{B})\boldsymbol{\eta} &= \boldsymbol{\Gamma\xi} + \boldsymbol{\delta\xi} \\
\boldsymbol{\eta} &= (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\Gamma\xi} + (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\delta} \\
\boldsymbol{\eta} &= \boldsymbol{\Pi_1\xi} + \boldsymbol{\Pi_2}
\end{aligned}$$

which can be interpreted as a linear model. Assuming normality within the model, the joint distribution of η and ξ is given by

$$f(z|\Omega)f(z_1, \ldots, z_N|\Omega).$$

with $\Omega \ni \Omega$ denotes the parameter space. Extracting one of the two elements in the partition, η and ξ, is done by considering the marginal distribution, given by

$$f(\eta, \xi|\Omega) = f(\eta|\xi, \Omega_1)f(\xi, \Omega_2)$$

where $\Omega_1 \in \Omega$ denotes the parameters for the conditional distribution of η given ξ and vice versa for Ω_2. The factorization displayed above assumes that the marginal distributions are given, or can be considered given.

As is the case within SEM, modeling typically revolves around the conditional distribution denoted above. Before finally introducing the notion of weak exogeneity, a technical definition should be discussed

Definition. (By Spanos, 1986). For any determined vector of parameters Ω_1, Ω_2 can take on any value within Ω.

Definition. (By Spanos (1986)) A variable ξ is weakly exogenous if and only if it is possible to construct a re-parameterization of Ω in terms of Ω_1 and Ω such that

- $\Omega = g(\Omega_1)$, or in other words, the respective parameters are a function of Ω_1
- Ω_1 and Ω_2 are *variation free*

Note that for the model described above, the parameters of interest are $\Omega^* = (\Pi_1, \Psi)$. The condition of ξ being weakly exogenous for this thesis, as

- The models used rely on the conditional estimation.
- More importantly, the violation of the exogeneity principle raises questions for the context of the model. Namely, the manipulation of the ξ variable by which we model the latent η are ill-considered whenever the parameters of ξ are a function of ξ.

4.1.3 Methods of assessment

As discussed in the previous section, the use of a χ^2-test is not without its controversy. Therefore, different statistical tests are considered for assessing the quality of the model. More specifically, the works of Hu and Bentler (1995) and Tucker and Lewis (1973) will be discussed in the paragraphs below.

Definition. The Comparative Fit Index is defined by

$$CFI = \frac{(\chi_b^2 - \mathrm{df}_b) - (\chi_q^2 - \mathrm{df}_q)}{\chi_b^2 - \mathrm{df}_b}$$

where χ_b^2 is the chi-square statistic of the so-called baseline model, or in other words the chi-square statistic of a model without any dependency (Hence with diagonal Ψ), and the χ_q^2 is the full model shown in figure 7.

- CFI $>$ .90 is considered an adequate fit (Hu and Bentler, 1995)
- CFI $>$ .95 is considered an excellent fit (Hu and Bentler, 1995)

Definition. The *Tucker-Lewis* index is defined as

$$TLI = \frac{\chi_b^2/\mathrm{df}_b - \chi_t^2/\mathrm{df}_t}{\chi_b^2/\mathrm{df}_b - 1}$$

Which shows a similarity with the CFI. Both are important examples of comparative fit indices, in which the model is compared to a baseline model, with maximum restrictions. It has the same guidelines,

- TLI $>$ .90 is considered an adequate fit
- TLI $>$ .95 is considered an excellent fit

Finally, the Root mean squared error of approximation is frequently used within structural equation models.

Definition. The *Root mean squared error of approximation*, or RMSEA, is defined as

$$\mathrm{RMSEA} = \sqrt{\frac{\chi^2 - \mathrm{df}}{\mathrm{df}(N-1)}}$$

and has the following guidelines (Brown and Cudeck, 1993)

- RMSEA $<$ .10 is considered an acceptable fit.
- RMSEA $<$ .08 is considered an adequate fit.
- RMSEA $<$ .06 is considered an excellent fit.

By Edward Rigon (1996), the CFI and TLI still have some problematic features. It is argued that the metrics are usefull in a exploratory setting (e.g. iteratively building up a model), and the RMSEA is aimed at evaluating a final and complete model. Therefore, this study will use the first two metrics when building the model, and will resort to the RMSEA after the final model has been constructed.

4.2 The structural model of music perception

Thus far, the mathematics underlying the structural equation model have been discussed. Theoretical notions of a path model and a measurement model formed the basis on which the models can be constructed. Now, it is time to construct the models evaluating all the previously listed principles. The first few models will be rather simple, and the models will get increasingly more complicated during this section. First, consider the simplest model as exemplified in section 4.1, modeling mood at the first musical piece and the second musical piece (time 1 and 2 respectively). Recall the equation (8), and define $y_1, \ldots, y_6$ the questions measuring mood at time 1 (e.g. the amount of happiness, sadness, ...). The items $y_7, \ldots, y_{12}$ are the same as $y_1, \ldots, y_6$, but then for mood at time 2. Hence, each y_i is the same question as y_{i+6}. Therefore, it is reasonable to let ε_i correlate with each ε_{i+6}. Additionally, let η_i designate the latent mood at time i, with $i = 1, 2$. The inclusion of the measurement of the latent mood at time 1 in the measurement of the latent mood at time 2 is also taken into account, as the mood at the second musical piece is intuitively interdependent upon the mood at the previous time. Displaying this construction visually,

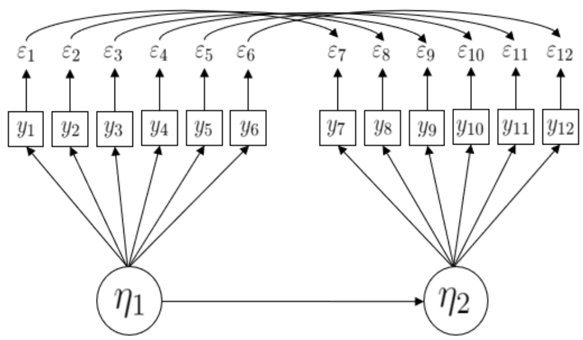

Figure 8: The simplest model.

A simple improvement to this model concerns the complete similarity between the measurements of η_1 and η_2. For each of the same questions, it is possible to impose a restriction on the factor loadings of $\mathbf{\Lambda}$ such that the factor loadings of y_i is equal to the factor loading of y_{i+6}. More precisely, each $\lambda_i = \lambda_{i+6}$.

The implementation of a complex structural model can be both time consuming and sensitive to errors. Therefore, in line with the previous function, it is also possible with the package coming along this thesis to automatically restrict the factor loadings of the items. The input requires a two dimensional vector with the two latent constructs (as exemplified above) in which the factor loadings of the items should be restricted. For example,

```
TidyLavaan::MeasurementModelConstraint(MetaData = meta_data,
SelectedFactors = c(mood1, mood2),
ConstraingLogical = TRUE ,
ConstraintLabel = 'b')
```

It is interesting to inspect the matrices underlying the model. For the model

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\Psi}$$

the matrices $\boldsymbol{B}$, $\boldsymbol{\Lambda}$ show

$$\boldsymbol{B} = \begin{bmatrix} 0 & 0.84 \\ 0 & 0 \end{bmatrix}, \qquad \boldsymbol{\Lambda} = \begin{bmatrix} 1 & 0 \\ 5.026 & 0 \\ 5.253 & 0 \\ 5.587 & 0 \\ 4.926 & 0 \\ 1.367 & 0 \\ 6.177 & 0 \\ 0 & 1 \\ 0 & 5.026 \\ 0 & 5.253 \\ 0 & 5.587 \\ 0 & 4.926 \\ 0 & 1.367 \\ 0 & 6.177 \end{bmatrix}$$

$$\boldsymbol{\Psi} = \begin{bmatrix} 0.96 & & & & & & & & & & & & & \\ 0 & 0.45 & & & & & & & & & & & & \\ 0 & 0 & 0.39 & & & & & & & & & & & \\ 0 & 0 & 0 & 0.35 & & & & & & & & & & \\ 0 & 0 & 0 & 0 & 0.45 & & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0.95 & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.12 & & & & & & & \\ 0.47 & 0 & 0 & 0 & 0 & 0 & 0 & 0.96 & & & & & & \\ 0 & 0.40 & 0 & 0 & 0 & 0 & 0 & 0 & 0.39 & & & & & \\ 0 & 0 & 0.43 & 0 & 0 & 0 & 0 & 0 & 0 & 0.33 & & & & \\ 0 & 0 & 0 & 0.57 & 0 & 0 & 0 & 0 & 0 & 0 & 0.34 & & & \\ 0 & 0 & 0 & 0 & 0.53 & 0 & 0 & 0 & 0 & 0 & 0 & 0.45 & & \\ 0 & 0 & 0 & 0 & 0 & 0.64 & 0 & 0 & 0 & 0 & 0 & 0 & 0.95 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.52 & 0 & 0 & 0 & 0 & 0 & 0 & 0.17 \end{bmatrix}$$

The factor loadings in $\boldsymbol{\Lambda}$ assigning a weight to each individual item is quite diverse. For a factor loading, a standardized factor loading > 0.5 is usually considered to be an adequate factor loading (Bagozzi & Yi, 1991). Then, after standardizing, for the items *Suprising, Relaxing, Emotional, Exciting, Beautiful* load very well onto the latent construct *mood*. On the contrary, the factor loadings of *Difficult* and *Demanding* indicate a poor loading onto *mood*. As these two factor are quite similar, the results seem intuitively cerebral. The ensure a suitable fit, the metrics show

Metric	**Value**
CFI	0.94
TLI	0.93
RMSEA	0.072

indicating an adequate fit. Now, the model discussed does, of course, not fully represent the entire data and hypothesized structure as we iteratively build up the model. Using the items measuring *meaning and interpretation* at, time 1 and 2, the model can be visualised by

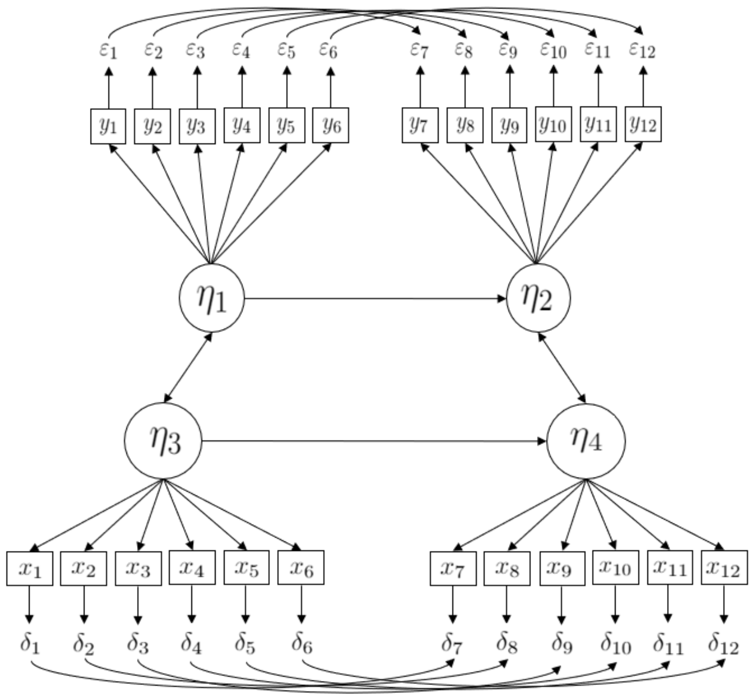

Figure 9: The second model

The fit indices are almost invariant under the addition of η_3 and η_4 to the model, as

Metric	**Value**
CFI	0.94
TLI	0.93
RMSEA	0.074

A simple inspection indicates that 24 of the 141 items have been covered in the model. The next addition comprises the mood of the participants at time 0 and 3. Denoting the mood at both times by η_5 and η_6 respectively, the model is updated into

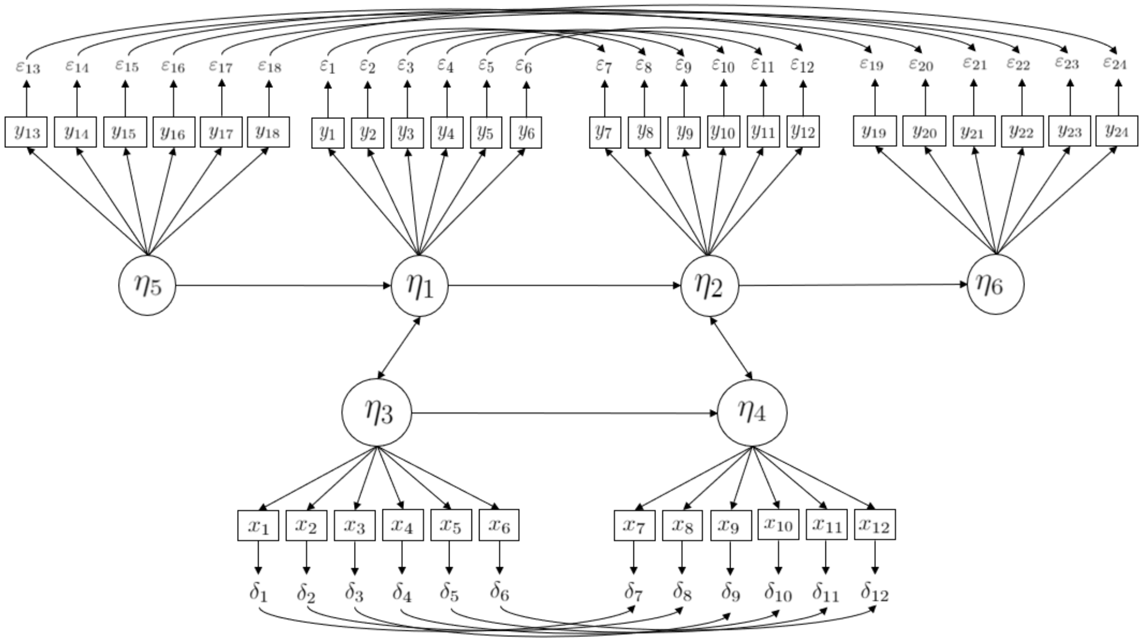

Figure 10: Third model

This time, the overall performance of the model drops to

Metric	**Value**
CFI	0.89
TLI	0.88
RMSEA	0.09

indicating a fit on the bounds of being a poor fit and an adequate fit. Since the inclusion of the latent construct shows a poor quality and the mood during the concert indicates a significantly better fit, these two variables will be dropped. Another possibility to tackle the latter issue is by internally improving the model. This sort of model improvement will only discussed after the final model has been established. In the final model, the exogenous variables are considered for implementation. However, the variables

1. *Setting*

2. *Reasons to visit the theatre*

3. *Quality of the musicians*

indicated a very poor fit, dropping all measures significantly (e.g. CFI = 0.69). This is due to a number of reasons, such as

- Different scales of measurement (ranging from binary data to 0-10 scaled variables) within the same latent construct

- A low variability within each item. This will be discussed in chapter five.

- Too much missing data, or data that is conditionally based on answering a different question (e.g. whether having a traffic jam whenever the method of transportation is a car).

The one latent construct within $\boldsymbol{\xi}$ improving the model is *Knowledge*, in which questions about the knowledge about the musical pieces are asked. The final model is therefore more complicated, and can be visualised by

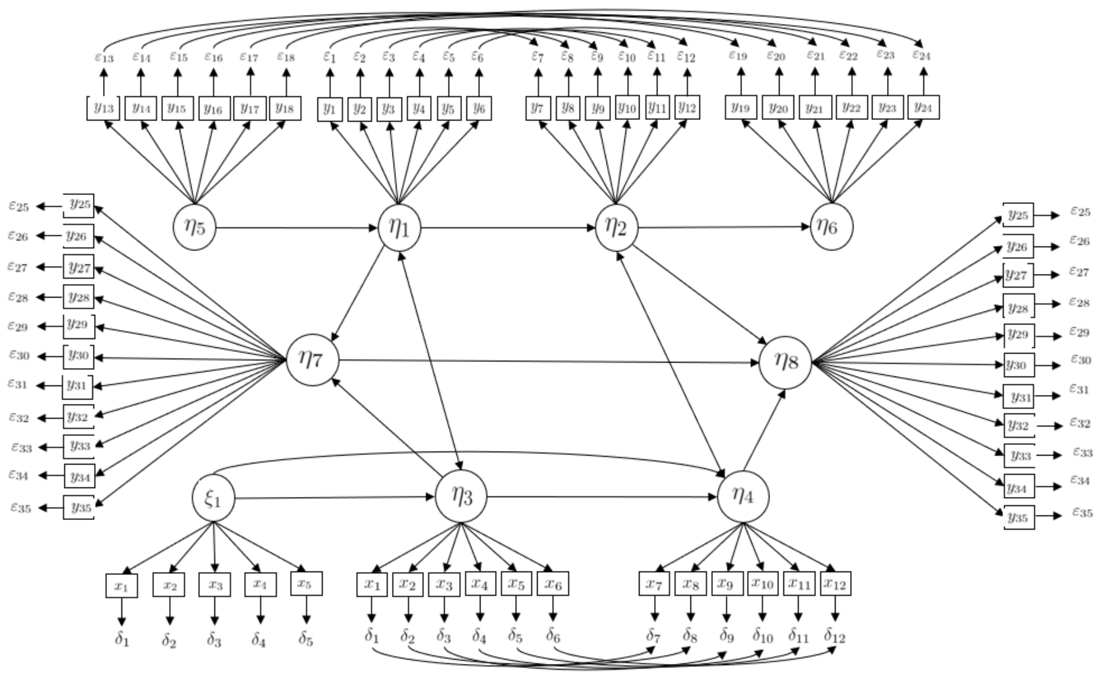

Figure 11: Full model

However, the fit indices have worsened. Although the RMSEA still shows the same adequate fit (0.074), the CFI has dropped to 0.85. Due to the inclusion of more factors while not considering the interaction between each item, the model has worsened in the fit indices. In section 4.2, the Modification Index has been established giving a powerful tool for studying plausible interactions within the model. For this study, the modification index is even more interesting as it shows which questions measure other latent variables. To prevent every possible modification to be included in the model, only improvements with $MI > 10$ (as is in line with most other statistical programs modeling modification indices, such as Mplus) will be considered and proposed to be added to the model. As is expected, most additions are rather obvious and consider correlation of items within the same latent factor. Such correlations improve the model without losing the hypothesized structure and are therefore convenient additions.

The inclusion of reasonable correlations can be done by the function *IncludeMI* in the package coming along this thesis. It can be executed by the code

```
TidyLavaan::IncludeMI(model)
```

To update the model, create a new model by the lines

```
NewModel <- paste(OldModel, '\n\n', TidyLavaan::IncludeMI(model))
```

and can be embedded in the lavaan/cfa function of the package lavaan.

A list of the additions and its considerations is given below. Note that only correlations are added. The modification index also indicates possible changes in the measurement model. However, to stay as close as possible to the original study, the addition of a statistically better measurement model will not be considered. It is also noteworthy that due to having a large and complex model, statistical significant modifications are easily obtained, although they might not be true in reality.

Item 1	Item 2	MI value	Considerations
I felt attracted to the developments within the musical piece (Piece 1)	I listened to music with my eyes closed (Piece 1)	10.68676	The MI-value is slightly above the considered norm. Since the correlation is not intuitively clear, this addition will not be considered
I felt attracted to the developments within the musical piece (Piece 1)	It moved me (Piece 1)	11.81207	Although the MI-value is slightly above the considered, the intuitive interdependence is worth taking into account and will be added to the model
I felt attracted to the developments within the musical piece (Piece 1)	It touched me (Piece 1)	10.48376	The MI-value is slightly above the considered norm. Since the correlation is not intuitively clear, this addition will not be considered
I felt attracted to the developments within the musical piece (Piece 1)	I like the music of Bach (Piece 1)	10.78627	Although the MI-value is slightly above the considered, the intuitive interdependence is worth taking into account and will be added to the model
Difficult (Piece 1)	Demanding (Piece 1)	32.8192	Intuitively, the correlation between the two items is quite clear. Also, the MI-value is highly significant.
Difficult (Piece 1)	It moved me (Piece 1)	10.59361	The MI-value is slightly above the considered norm. Since the correlation is not intuitively clear, this addition will not be considered
Relaxed (Piece 1)	Exciting (Piece 1)	14.2452	The correlation between the two items this time is reverse coded, as the can be considered each other opposites.
Relaxed (Piece 1)	Gave me inner rest (Piece 1)	15.67119	The MI-value is slightly above the considered norm. Since the correlation is not intuitively clear, this addition will not be considered
Relaxed (Piece 1)	Emotional (Piece 1)	11.22442	The correlation between the two items this time is reverse coded, as the can be considered each other opposites
It Touched me (Piece 1)	It moved me (Piece 1)	10.65942	Although the MI-value is slightly above the considered, the intuitive interdependence is worth taking into account and will be added to the model
It touched me (Piece 1)	Emotional (Piece 1)	15.56084	Although the MI-value is slightly above the considered, the intuitive interdependence is worth taking into account and will be added to the model
It touched me (Piece 1)	Demanding (Piece 1)	10.50695	The MI-value is slightly above the considered norm. Since the correlation is not intuitively clear, this addition will not be considered
It touched me (Piece 1)	Gave me inner rest (Piece 1)	11.94732	The MI-value is slightly above the considered norm. Since the correlation is not intuitively clear, this addition will not be considered
Difficult (piece 2)	Demanding (piece 2)	61.5717	Intuitively, the correlation between the two items is quite clear. Also, the MI-value is highly significant.
Provided consolation (piece 2)	Made me feel compassion (piece 2)	66.32917	Intuitively, the correlation between the two items is quite clear. Also, the MI-value is highly significant.

Figure 12: Modification index. Actual table is triple the size

After the implementation, the model has improved, as

Metric	Value
CFI	0.90
TLI	0.90
RMSEA	0.053

indicating an adequate to good fit. As for the confirmatory model the RMSEA is the more important metric and shows an excellent fit. Also note that the Hessian is positive definite and that the degrees of freedom (df = 1865) is positive - resulting in identification.

4.3 Regression coefficients and variable selection

The assessment of the quality of the model can be chosen in two separate ways. First of all, constructing the method by adding and dropping variables purely based on model performance is an option. Secondly, by staying close to the data, the point of view is not statistical performance, but the performance of the data in a model-theoretic context. Thus far in this thesis, a combination of both methods have been used. Although the writer is not a psychologist nor a sociologist, by means of educated guesses a statistical valid model was constructed. Variables which indicated significant model performance loss were dropped from the model. To ensure a valid representation of the data, only factors were dropped - and not individual items. This section assesses the output of the structural equation model, while specifically assessing the quality of the measurement model.

The concept of regression in a structural equation model is similar to that of a linear regression. Note that the regression coefficients can be found in the $\boldsymbol{B}$-matrix of the structural equation. Extracting the elements from the $\boldsymbol{B}$-matrix, the non-zero elements show

Latent Variable	**Correlates with**	**Standard Error**	**Estimate**
Meaning and Interpretation t_1	Knowledge	0.045	0.280
Meaning and Interpretation t_2	Meaning and Interpretation t_1	0.037	0.840
Meaning and Interpretation t_2	Knowledge	0.81	0.280 *
Mood t_2	Mood t_1	0.036	0.810
Experience t_1	Mood t_1	0.683	-2.407
Experience t_1	Meaning and Interpretation t_1	0.095	0.379
Experience t_2	Experience t_1	0.060	0.075*
Experience t_2	Mood t_2	0.788	-3.348
Experience t_2	Meaning and Interpretation t_2	0.071	0.293

Table 2: Results of latent factor score improvement per partition of knowledge

As *Knowledge* presumably exerts influence on the score of the construct of *meaning an interpretation*, the analysis shows that as *Knowledge* is already incorporated in the regression of *meaning and interpretation* at time 1, there is no need to let the *meaning and integration* at t_2 also model knowledge. Including solely the regressions which are significantly different from zero, the new regression coefficients are similar to the previous table, as

Latent Variable	**Correlates with**	**Standard Error**	**Estimate**
Meaning and Interpretation t_1	Knowledge	0.045	0.289
Meaning and Interpretation t_2	Meaning and Interpretation t_1	0.037	0.876
Mood t_2	Mood t_1	0.037	0.816
Experience t_1	Mood t_1	0.682	-2.334
Experience t_1	Meaning and Interpretation t_1	0.095	0.382
Experience t_2	Mood t_2	0.788	-3.411
Experience t_2	Meaning and Interpretation t_2	0.071	0.330

Table 3: Results of latent factor score improvement per partition of knowledge

The interpretation of the regression results are similar to a normal linear regression. Since all items and latent variables operate on an uniform scale, it can be obtained that *mood* quite heavily impacts the experience at both times. Maybe somewhat surprising, the opposite is true for the *meaning and interpretation*. At both times, the regression coefficients do not differ significantly from zero in the regression model. Secondly, it is noteworthy to discuss the individual factor loadings. The construction of the items loading onto the factors is based on sociological and psychological knowledge. By using mathematics, or more specifically the constructed structural equation model, it is possible to identify whether the latent constructs are measured correctly. If the measurement estimate is not significantly different from zero, the consideration of removing the item is suggested. Then, the inspection of the measurement model is required. The $\boldsymbol{\Lambda}$ matrix identifying the factor loadings has been converted into a readable table and is listed below.

Latent variable	**Item**	**Coefficient**
Meaning and Interpretation t_1	I recognized the emotions	1
Meaning and Interpretation t_1	I felt attracted by the developments	1,122
Meaning and Interpretation t_1	I understood the intentions of the composeer	0,925
Meaning and Interpretation t_1	I tended to physically move	0,654
Meaning and Interpretation t_1	I primarily listened with my eyes closed	0,335*

Meaning and Interpretation t_2	I recognized the emotions	1
Meaning and Interpretation t_2	I felt attracted by the developments	1,122
Meaning and Interpretation t_2	I understood the intentions of the composer	0,925
Meaning and Interpretation t_2	I tended to physically move	0,654
Meaning and Interpretation t_2	I primarily listened with my eyes closed	0,335*
Mood t_1	difficult	1
Mood t_1	surprising	2,238
Mood t_1	relaxing	1,863
Mood t_1	emotional	3,296
Mood t_1	exciting	2,687
Mood t_1	Demanding	$-0,378^*$
Mood t_1	beautiful	2,538
Mood t_2	difficult	1
Mood t_2	surprising	2,238
Mood t_2	relaxing	1,863
Mood t_2	emotional	3,296
Mood t_2	exciting	2,687
Mood t_2	Demanding	$-0,378^*$
Mood t_2	beautiful	2,538
Experience t_1	Gave me a feeling of beauty	1
Experience t_1	Touched me	1,197
Experience t_1	Appealed to my fantasy	0,865
Experience t_1	challenged my feeling for music	1,03
Experience t_1	Provided inner rest	1,065
Experience t_1	Made me experience a new part of the piece	0,981
Experience t_1	Provided consolation	1,067
Experience t_1	Made me feel sympathy	0,957
Experience t_1	Made me lose track of time	1,089
Experience t_1	Deeply impressed me	1,298
Experience t_1	Made me able to clear my head	1,051
Experience t_2	Gave me a feeling of beauty	1
Experience t_2	Touched me	1,197
Experience t_2	Appealed to my fantasy	0,865
Experience t_2	challenged my feeling for music	1,03
Experience t_2	Provided inner rest	1,065
Experience t_2	Made me experience a new part of the piece	0,981
Experience t_2	Provided consolation	1,067
Experience t_2	Made me feel sympathy	0,957
Experience t_2	Made me lose track of time	1,089
Experience t_2	Deeply impressed me	1,298
Experience t_2	Made me able to clear my head	1,051
Knowledge	I know the music of Beethoven quite well	1
Knowledge	I know the music of Bach quite well	0,924
Knowledge	I like the music of Bach	0,754
Knowledge	I know the music of Brahms quite well	0,979
Knowledge	I like the music of Brahms quite well	0,592

Merely the items

- *I primarily listened to the music with my eyes close*
- *It was quite a demanding musical piece*

did not fit the model well, also due to a relative high standard error. As the distinction between being significantly different from zero is set by some crude interval, usually .95, some comments about other variables should be made. Then, the items

- *I tended to physically move*
- *Appealed to my fantasy*
- *I like the music of Brahms quite well*

did not seem to fit the model too well either. In the next chapter, *item response theory* will be discussed. By using the *discriminability parameter* within such a model, it is possible to extract how much utility a question has to offer to distinguish a participant based on the score. The 'future' of the variables just listed will be assessed by the amount of discriminability they have to offer. Only after the latter assessment, a new structural equation model will be constructed such that a more optimized model can be formed based on diverse and informed considerations.

4.3.1 An alternative method of locating goodness of fit

Locating the residuals underlying the model, the difference between the implied and the population covariance matrix can be computed. The residuals of $|S - \Sigma|$ can identify the places where the fit could be improved. Computation yields a lower diagonal matrix. Since we work with many observations, a simple matrix does not fulfill the needs of identifying a pattern within the residuals. This problem is analogous to the ones of identifying correlations within a large data set, which is usually resolved by plotting the correlation plot. The same principle is executed, with the only difference that the loadings, or values, are based on the residuals rather than the correlations.

The alteration of the correlation plot is made possible by including the statement

```
corplot = FALSE
```

in the corrplot function in the stats package. Then, by entering the matrix of residuals, the plot below can be constructed

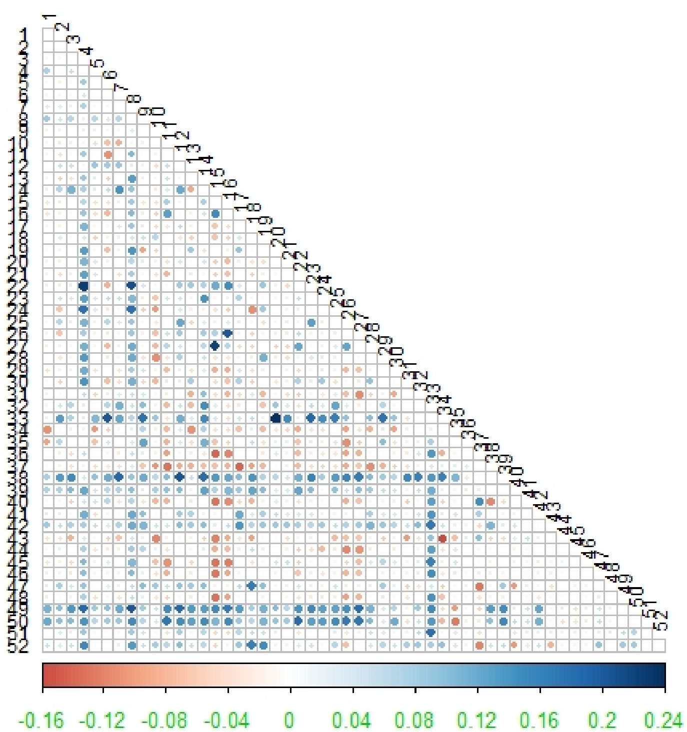

Figure 13: Plot of residuals as it was a correlation

Variables 49 and 50 (*tended to physically move* and *I listened to the music primarily with my eyes cloded*) show an interesting pattern of having consistently high residuals. Although these items could be omitted from the data, the hypothesized structure is changed within each factor, which declines the resemblance of the data.

As an adequate fit has been established, the next step in constructing analysis on the latent perception of music can be found both in a structural equation model as well as in other mathematical theories. However, within a structural equation model, it should be noted that latent factor estimation is not without problems. Analysing the hypothesized perception further, the primary objective is to identify the possible changes in the perception of music over time. The method of a latent growth model is frequently referred to as modeling the time component within a structural equation model. As the latent growth model is based on the same principles as a normal structural equation model, the same problems with latent inferences may occur. Also, the literature of a latent growth model is quite scarce, and therefore it relies on the framework built by Bollen.

4.3.2 Small recap

The original study compromised 141 questions ranging over eleven factors. Unfortunately, not all questions were did fit the model well. The items in the structural model described in the previous sections do describe the data well, and the factors

Factor	Time	Inclusion
Reasons of visiting the theatre	0	No.
Knowledge about the composers	0	Yes
Setting of the concert	all	No. Measured well, but no good addition to the model
Meaning and Interpretation	1,2	Yes
Mood	1,2	Yes
Experience	1,2	Yes
Break	1,2	No. Not well-defined nor numerically valid to include
Musical background	all	No. Numerical Problems
Quality of the musicians	all	No. Little to no variability

As stated in the table above, the quality of musicians is not taken into account for this model. Although the measure was initially to be included in the model, the model performance dropped. Inspection shows that the variability within the measures is very low. It should be noted that the exclusion of the quality of musicians does not imply the quality of musicians itself was poor. On the contrary, almost every participant rated the musicians very high, hence indicated a good quality but points out the lack of providing information for the model.

4.4 Factor Estimation

Constructing and assessing the structure of latent variables is often a goal in scientific studies. By means of measurement models as introduced above, the latent variables are given a more concrete and practical form, although a numerical value is lacking. The latter numerical interpretation of a latent variable is usually referred to as a *factor score*, and numerous factor score estimation procedures have been developed over the latter centuries. Having alternative methods to examine factor scores is, however, not necessarily desired. Since most methods are in line with the theory of factor analysis, and in our case a structural model, it poses different well-accepted options to assign different numerical values to a latent variable. The paradoxical situation in which multiple estimators are well-defined, but are unequal to each other, is called *factor indeterminacy.* To align with the most common methods applied to a structural equation model, the regression method and the Bartlett method will be introduced first. Then, a more recent alternative will be discussed.

$$\xi = \boldsymbol{A^T x}.$$

However, the factor estimation procedure should preserve the structural model, with

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda \Phi \Lambda^T} + \boldsymbol{\Psi}$$

with Σ being the covariance matrix of x, $\boldsymbol{\Phi} = \mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^T]$ is the common factors correlation matrix and $\boldsymbol{\Psi} = \mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T]$. The factor score estimation within the software used for this thesis, Lavaan, is *Bartlett's method.*

Definition. In *Bartlett factor scores* only the shared factors influence the factor scores. The summation of the squared components of the unique factors sequencing along the variables in minimized. Then, the matrix A^T is computed by

$$\boldsymbol{A}^T = (\boldsymbol{\Lambda}^T(\boldsymbol{\Psi}^{-1})^T\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}^T\boldsymbol{\Psi}^{-1} \tag{9}$$

The matrix $\boldsymbol{A}^T$ is constructed by means of a maximum likelihood estimate, resulting in unbiased estimates of the factor scores (Herschberger, 2005). Nevertheless, this method poses not merely positive sides to factor score estimation, as the method does not take the structural matrix $\boldsymbol{\Phi}$ into account. The notion of *correlation preserving* methods tries to formalize this concept

Definition. Let A be the $m \times q$ matrix representing the matrix of coefficients in the factor score equation $\boldsymbol{\xi} = \boldsymbol{A}^T\boldsymbol{x}$. An estimator of $\boldsymbol{A}^T$ is correlation preserving if and only if

$$\mathbb{E}[\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}^T] = \mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^T] = \boldsymbol{\Phi}$$

is satisfied.

Lemma. The matrix $\boldsymbol{A}$ as discussed before is of the form

$$\boldsymbol{A} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{C}\boldsymbol{\Omega}^{1/2}$$

with $\boldsymbol{C}$ as some matrix with dimension $m \times q$ (Ten Berge, Krijnen, 1999). The proof follows trivially from the inversion the structural decomposition of $\boldsymbol{\Sigma}$

Theorem. For the estimator $\hat{\boldsymbol{\xi}}$, the matrix $\boldsymbol{C}$ must be columnwise orthonormal, i.e. $\boldsymbol{C^T C} = \boldsymbol{I}$ to preserve the correlation.

Proof: Using the previous lemma, a simple computation shows

$$\begin{aligned}\boldsymbol{A} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{C}\boldsymbol{\Omega}^{1/2} \implies \mathbb{E}[\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}^{\mathbb{T}}] &= \mathbb{E}[\boldsymbol{A}^T\boldsymbol{x}(\boldsymbol{A}^T\boldsymbol{x})^T] \\ &= \boldsymbol{A}^T\mathbb{E}[x\boldsymbol{x}\boldsymbol{x}^T]\boldsymbol{A} \\ &= \boldsymbol{A}^T\boldsymbol{\Sigma}\boldsymbol{A} \\ &= (\boldsymbol{\Omega}^{1/2})^T\boldsymbol{C}^T(\boldsymbol{\Sigma}^{-1/2})^T\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{C}\boldsymbol{\Omega}^{1/2}\end{aligned}$$

Note that $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ are symmetric, thus

$$\begin{aligned}&= (\boldsymbol{\Omega}^{1/2})\boldsymbol{C}^T\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{C}\boldsymbol{\Omega}^{1/2} \\ &= \boldsymbol{\Omega}^{1/2}\boldsymbol{C}^T\boldsymbol{I}\boldsymbol{C}\boldsymbol{\Omega}^{1/2} \\ &= \boldsymbol{\Omega} \iff \boldsymbol{C}^T\boldsymbol{C} = \boldsymbol{I}\end{aligned}$$

It can be shown that the Bartlett estimator is not correlation preserving by checking that $\mathbb{E}[\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}^T] \neq \boldsymbol{\Phi}$. In the latter century, Anderson and Rubin proposed the first correlation preserving method by means of minimizing the weighted least-squares function $f(\hat{\boldsymbol{\xi}})$ subjected to the previous lemma. However, this solution only considered $\boldsymbol{\Omega} = \boldsymbol{I}$. MacDonald (1981) generalized the concept and extended the method such that $\boldsymbol{\Omega}$ could be taken arbitrarily. More precisely,

$$\boldsymbol{A} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{C}\boldsymbol{\Omega}^{1/2}$$

where $\boldsymbol{C}$ is determined by the singular value decomposition

$$\boldsymbol{U}\boldsymbol{\Delta}\boldsymbol{V}^T = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}\boldsymbol{\Omega}^{1/2}$$

with $\boldsymbol{U}$ and $\boldsymbol{V}$ as orthogonal matrices and $\boldsymbol{\Delta}$ a diagonal matrix. Green proposed a different method in which the function

$$g(\boldsymbol{A}) = \operatorname{trace}\operatorname{MSE}(\boldsymbol{A}^T\boldsymbol{x}) \quad \text{subjected to} \quad \boldsymbol{A} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{C}\boldsymbol{\Omega}^{1/2}$$

is minimized. Or, in other words, the of the trace of the mean squared error is minimized. Similarly to Anderson and Rubin, the solution relied upon the singular value decomposition with

$$\boldsymbol{U}\boldsymbol{\Delta}\boldsymbol{V}^{\boldsymbol{T}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Omega}^{3/2}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are again orthogonal matrices and Δ is a diagonal matrix. Krijnen et al (1996) generalized the two notions above by means of specifying an iterated method for determining $\boldsymbol{C}$ in which the determinant of the mean squared error matrix is minimized. Ten Berge, Krijnen (1999) improved this method by considered a solution of $\boldsymbol{C}$ in a closed form, and the iterative procedures are no longer required. As shown in Ten Berge, Krijnen (1999),

$$\boldsymbol{C} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{L}(\boldsymbol{L}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{L})^{1/2}$$

with $\boldsymbol{L} = \boldsymbol{\Lambda}\boldsymbol{\Omega}^{1/2}$.

Corollary. The factor score estimation procedure proposed by Ten Berge and Krijnen is correlation preserving.

Proof: Referring to the previously constructed lemma in which $\boldsymbol{C}^T\boldsymbol{C} = \boldsymbol{C}\boldsymbol{C}^T = \boldsymbol{I}$ is a sufficient criteria for a correlation preserving method, a computation shows

$$\begin{aligned}
\boldsymbol{C}\boldsymbol{C}^T &= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{L}(\boldsymbol{L}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{L})^{1/2}\big\{\boldsymbol{\Sigma}^{-1/2}\boldsymbol{L}(\boldsymbol{L}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{L})^{-1/2}\big\}^T \\
&= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{L}(\boldsymbol{L}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{L})^{-1/2}((\boldsymbol{L}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{L})^{1/2})^T\boldsymbol{L}^T\boldsymbol{\Sigma}^{-1/2} \\
&= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{L}(\boldsymbol{L}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{L})^{-1}\boldsymbol{L}^T\boldsymbol{\Sigma}^{-1/2} \\
&= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{L}\boldsymbol{L}^{-1}\boldsymbol{\Sigma}(\boldsymbol{L}^T)^{-1}\boldsymbol{L}^T\boldsymbol{\Sigma}^{-1/2} \\
&= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2} \\
&= \boldsymbol{I}
\end{aligned}$$

which completes the proof. Hence, the method proposed by Ten Berge and Krijnen consists of the most general method while preserving correlation.

This method has been implemented in the TidyLavaan package coming along this thesis. It can be used by executing the lines

```
TidyLavaan::TenBergeKrijnen(LavaanModel)
```

The only input required is the Lavaan model. The software extracts the matrices Σ, L, Ω, Φ and Ψ and the implemented data for the user. If new data is required, the following code should be executed

```
TidyLavaan::TenBergeKrijnen(LavaanModel, NewData)
```

4.4.1 Results

As a thorough theoretical foundation has been established, the application of the theory illustrates the capacity of the theory itself.

The results discussed in this sections are based on the factor scores of the Mood of a participant at entrance time and at the time after the concert. Hence, the difference in Mood when entering the theatre and leaving the theatre after the concert is assessed. It should be noted that Mood has also been measured at both classical pieces displayed in the theatre. However, for these specific times, a more nuanced method of measuring Mood is required. Globally, the method of factor estimation as discussed in the latter section suffices, but when studying the Mood (or any other latent factor) requires more nuance in the mathematics, which will be discussed in the upcoming chapter by means of item response theory.

After the implementation of the methods of the previous section in R, the factor scores based on the most complete model show (only the first few observations are taken)

Experience t_1	Mood t_1	Mood t_2	Interpretation t_1	Interpretation t_2	Experience t_2	Knowledge
1,45	-1,06	-0,58	-1,2	-0,61	-0,84	1,18
-0,76	0,45	0,35	-1,77	-0,67	0,53	-0,79
0,37	-0,15	0,08	-0,56	-0,8	0,3	-0,39
-1,01	-0,32	0,08	0,09	-0,02	-0,07	-1,35
0,38	0,12	-0,07	0,98	0,61	0,37	1,85
0,24	-0,17	0,22	0,1	0,2	-0,06	-0,17

By construction, $\mathbb{E}[\xi] = 0$. Checking this construction yields

Latent Factor ξ	$\mathbb{E}[\xi]$
Mood t_0	4.775892e-17
Mood t_1	-1.513396e-16
Mood t_2	1.024873e-16
Interpretation t_1	1.024428e-16
Interpretation t_2	2.794344e-17
Mood t_4	-4.186291e-17
Knowledge	8.162729e-17

which is in line with the construction of the model. Visually, the distribution

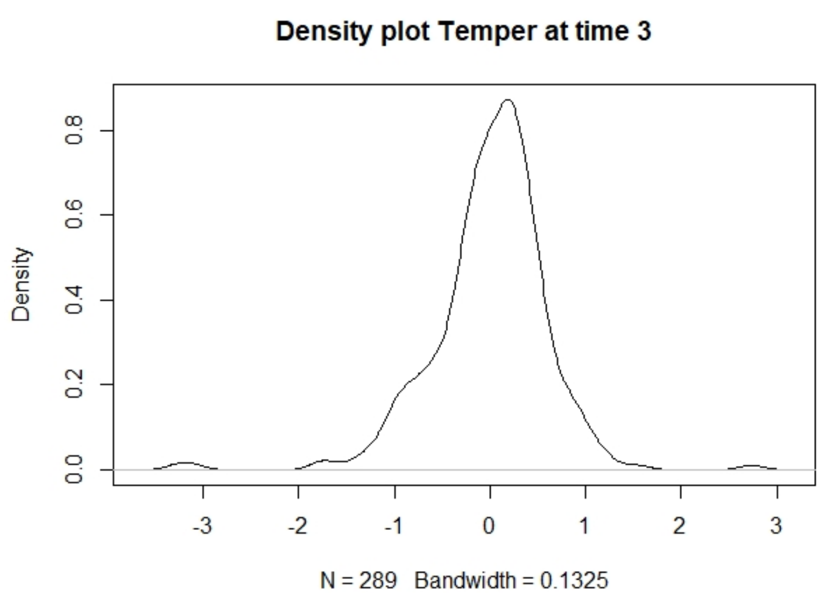

Figure 14: Density plot of the factor estimation

By a transformation into the scale of the measurements underlying the latent factor, a better interpretable estimation is obtained

The rescaling has been implemented in the TidyLavaan package coming along this thesis. It can be used by executing the lines

```
TidyLavaan::FactorRescale(FactorEstiamtes)
```

The only input required is the matrix of factor estimates. It automatically rescales the factor estimates into the known scale of the measurements underlying the model. If different measurement scales are used, please pass it through the function by the additional argument *scale* or use the rescale function form the tidyr-package

```
TidyLavaan::TenBergeKrijnen(LavaanModel, scale = c(a,b))
```

After a rescaling, the first few lines of the data shows

Experience t_1	Mood t_1	Mood t_2	Interpretation t_1	Interpretation t_2	Experience t_2	Knowledge
3,42	3,53	4,9	1,15	2,35	3,81	2,91
4,37	2,18	4,36	2,12	2,09	3,58	3,26
3,2	1,81	4,36	2,64	3,59	3,21	2,43
4,37	2,79	4,05	3,35	4,79	3,65	5,22
4,26	2,15	4,64	2,65	4	3,23	3,46
5,1	2,99	3,66	0,87	5,84	3,38	2,58

Asserting changes over time can be done by means of constructing a confidence interval for each individual. The confidence interval is of the form

$$[x - z_\alpha \text{s.e.},\ x + z_\alpha \text{s.e.}]$$

where s.e. denotes the standard error, defined by

$$\text{s.e.} = \frac{\sigma}{\sqrt{n}},$$

z denotes the respective z-score and α denotes the desired confidence value, which will be set to 0.95. Then, comparing two similar factor scores at different times can be assessed by checking whether the factor score of the factor score at time 2 is contained in the confidence interval of the factor score at time 1.

The construction of the confidence interval is contained in the package coming along this thesis and can be called upon by the line

```
TidyLavaan::DetectChanges(Column1, Column2)
```

The output is a data frame, which the two respective columns, a lower and an upper bound for the confidence interval, the dichotomous variable *Significant Change* detecting whether the value in column 2 is contained in the confidence interval of column 1, and an extra column checking whether the factor score has become bigger or smaller.

Based on the function described above, the results will be summarized in the final subsection of this chapter. Reassuring enough, the data and models show that the participants' Mood improved in about 80% of the cases. More precisely

Improved	Observations
No	54
Yes	235

Table 5: Results of latent factor scores

Also, studying the amount of improvement of the factor score of Mood can be computed. This can, again, be done by partitioning the participants in groups showing improvement in Mood or no improvement.

Improved	Average Improvement
No	-20.6 %
Yes	55.8 %

Table 6: Results of latent factor score improvement

Thus, on average, people who show a decrease in the factor score of Mood show a 20% decline. On the upside, participants showing an increase in the factor score related to Mood show a 55% increase in the factor score.

Interestingly enough, when constructing a partition in the data based upon the factor score of the knowledge, the data shows a very close similarity for the group having an insufficient knowledge of classical music and a sufficient knowledge of classical music. Concentrating on the 25% quantile performing the best on knowledge, the improvement in Mood for both the group is almost completely similar:

Knowledge score	Improvement	Observations (in %)
0-75% of knowledge scores	No	16.3%
0-75% of knowledge scores	Yes	83.7%
75-100% of knowledge scores	No	16.2%
75-100% of knowledge scores	Yes	83.8 %

Table 7: Results of latent factor score improvement per partition of knowledge

It might be interesting to combine the latent variables into one plot in order to notice patterns. As the regression coefficient of knowledge indicates a strong relationship, it is noteworthy to discuss the interdependence of experience and knowledge. By plotting a simple visual with all the knowledge scores on the x-axis and the average factor score of experience at both times, on the y-axis, we have

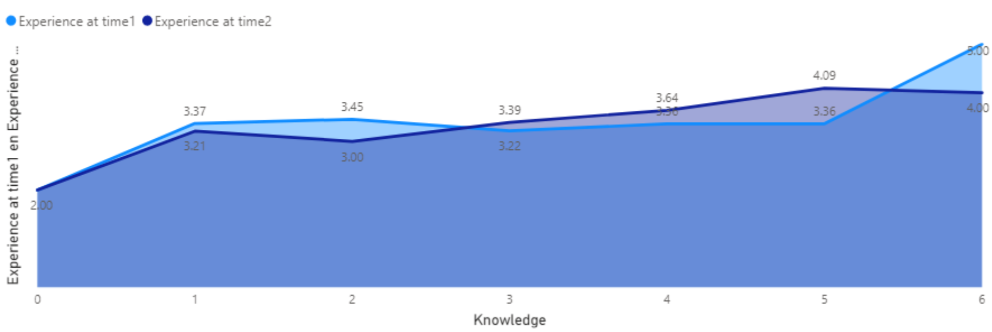

Figure 15: Visualiation of the interaction of knowledge and experience

showing an intuitively pleasing relationship with knowledge and the experience. A different effect can be obtained for *meaning and interpretation*

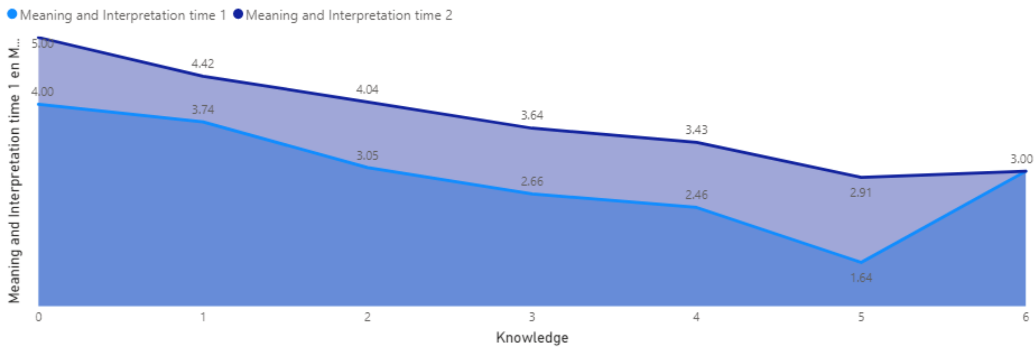

Figure 16: *Meaning and Interpretation per knowledge score*

Hence, obtaining a higher score on the knowledge axis results in having a lower score at the factor of *meaning and interpretation*. Psychologically, this result may arise from the paradoxical situation in which an informed person is more (self-) critical than a worse informed person. However, such a study is beyond the scope of this thesis and acts merely an educated guess by the author. Finally, a more understandable and clear pattern shows when *mood* is used as the x-axis,

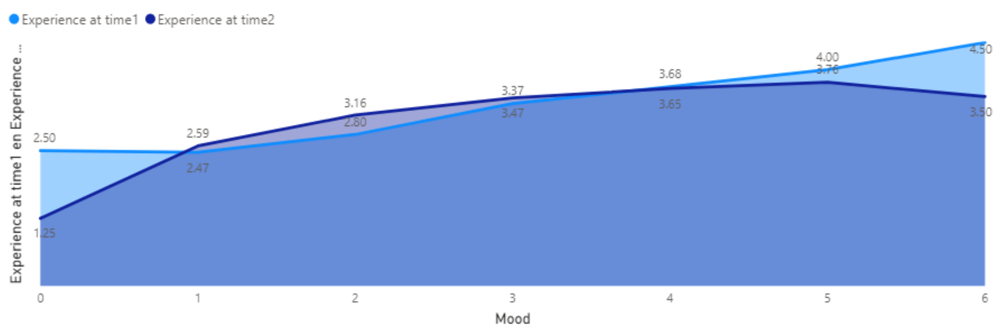

Figure 17: Visualisation of the interaction of *mood* and *experience*

The experience gradually improves with mood, which is in line what is intuitively to be expected.

4.5 Conclusion and discussion

This chapter presented a straight forward method for constructing the structural equation model. By means of a combination of educated guesses, proper scientific conduct and statistical performance, the model has been generated iteratively. It should be noted that model performance could have been much higher if it were not for the inclusion of seemingly important variables, such as mood and experience. As the goal of building the model is not statistical significance, but finding a method to study, investigate and test a hypothesized or pre-defined knowledge, model performances dropped over the iterations. The measurement model plays an important role in the statistical power and therefore the appropriateness of the loadings in $\mathbf{\Lambda}$ has been discussed. Since a few items are on the edge of statistical significance and such an edge is somewhat arbitrary, a different mechanism is required to get more insight into the quality of the factors. The proposed method is the *item response theory*, which brings parameters such as *discriminability* into consideration.

5 Latent Inferences and Measurement

The first and foremost method of research with survey is measurement. Analysing financial insecurity, academic achievement or in this case the perception of music is performed by conducting measurements, partitioned into different categories and questions. The theory of conducting a proper survey heavily relies upon the qualities of the questions, the phrasing of the questions, the possible bias of the research and the participants of the survey. For instance, the legitimacy of a poll fully depends upon the quality and how well the sample represents the full society (e.g. consider the polls of the 2016 presidential election in the USA). The central topic of this section comprises the theory for measurement.

5.1 Item Response Theory

The Item Response theory encompasses a number of mathematical models, including the possibility of statistical inferences focused on addressing and assessing questions from a survey, or **items** as they are called within this theory. Analysis of items, assessing scores and latent variable analysis are all part of the latter theory, promising a valid method for research.

Using a theory or statistical method arbitrarily cannot be categorized as conducting proper science. Therefore, a case must be made for using item response theory. Traditional statistics frequently resorts to using the mean statistics for comparing the overall score on the same item at two different times. However, this sort of simplification of the problem is not possible for this model. Each latent category of questions contains multiple items. Therefore, comparing the total score of each category would implicate using the mean of the mean. This kind of a statistic is necessarily a valid statistic for displaying the entire data. The concern made above can be exemplified by considering the sample questions.

The items for the survey at hand admits an ordinal scheme on a 0-6 scale. The scale can be interpreted as a **Likert scale**, based on the interpretation of ranging from Very Strongly disagree to Very Strongly Agree, with neutral as the median possible score. However, classical item models usually depend upon *dichotomy*, i.e. a binary scale. The section below builds up a item response theory model, starting by the most simplest binary data.

5.2 The Rasch model

Developed by Georg Rasch, the *Rasch model* is a method for modeling dichotomous data. The model compromises the trade-off between the participants knowledge and/or capabilities and the *'difficulty'* of a question. More precisely, let y_{ij} be the jth item of the questionnaire, and let i be the ith response such that the questionnaires has dimensions $i \times j$. The items measure the dichotomous latent variable *ability* θ, which can be interpreted as knowledge, enjoyment or capability dependent upon the context. The equation providing the base for the Rasch model is the Sigmoid function $S(x)$ for $x = \theta_i - b_j$

$$\mathbb{P}(Y_{ij} = y_{ij}|\theta_i) = \frac{e^{\theta_i - b_j}}{1 + e^{\theta_i - b_j}}$$

where b_j is the item *difficulty*, or the amount of difficulty one has with getting the score '1' on this question. To yield valid results, proper statistical assumptions should be imposed. More specifically,

1. **Monotonicity**. Similar to a cdf, the Probability of the Rasch model is strictly increasing, with $0 \leq \mathbb{P}(y_{ij} = 1|\theta_i) \leq 1$. More formally, for $\theta_i > \theta_m$:

$$f(y_{ij}|\theta_i, b_j) > f(x_{im}|\theta_m, b_j), \quad \forall \theta_i, \theta_m$$

2. **Unidimensionality** The respective items are restricted to interact with only one latent trait. Therefore, the vector form of θ_i is strictly one dimensional. Mathematically,

$$\mathbb{P}(y_{ij} = 1|\theta_i, b_j, \varphi) = \mathbb{P}(y_{ij} = 1|\theta_i, b_j)$$

3. **Conditional independence.** The items can only be interdependent because of the depedence on the latent variable θ_i. Thus,

$$\mathbb{P}(Y_{i1} = y_{i1}, \dots, Y_{iJ} = y_{iJ} | \xi_i) = \prod_j f_j(\xi_j)$$

4. **Sufficiency.** The raw score $r_i = \sum_i y_{ij}$ has all information about the ability. By the factorisation theorem,

$$f(y_{ij}, \dots, y_{kj} | \theta_i) = g(r_i | \theta_i) h(x_{i1}, \dots, x_{iJ})$$

For each of the assumptions listed above, a more detailed explanation is required. Also, algorithms to check the conditions are shared.

5.2.1 A likelihood ratio model for assessment

Several tests have been constructed to check assumption (1) and (4), such as Molenaar's Statistics U and Fisher and Scheiblechner's statistic S. The most common statistical test is constructed by Anderson, who also formulated the theory of polytomous Rasch models, which will be discussed later in this chapter. For the Andersen test, the conditional likelihood, given in Reinhold (2018), is defined by

$$L = \frac{\exp(-\sum_j b_j r_i)}{\prod_r \left(\sum_y \exp(-\sum_j y_j b_j) \right)^{n_r}}$$

The likelihood can be partitioned into a likelihood estimate for each possible raw score such that

$$L = \prod_r L^{(r)}$$

Then, by a simple deconstruction it follows that

$$L^{(r)} = \frac{\exp(-\sum_j b_j r_i)}{\left(\sum_y \exp(-\sum_j y_j b_j) \right)^{n_r}}$$

if the raw score r_i is a sufficient statistic, the two likelihood should be similar. However, if the opposite is true, then the raw score is not a sufficient statistic and hence the fourth condition is violated. Also, if the *item characteristic curve* for a question j has a non-monotonic property, then $|\theta_i|$ varies along the different sub scores. Then, the the following LR-test Λ will decrease. By Anderson (1973), the statistic

$$Z = -2 \log \Lambda = 2 \sum_r L^{(r)} - 2 \log L$$

approximately follows has the $\chi^2_{(j-1)(j-2)}$ distribution (Andersen, 1973b). The algorithm checking the statements above can be constructed using

Algorithm 2 Andersen LRT

```
procedure AndersenLRT(data)                                  ▷ Clean data
    L ← exp(−∑_j b_j r_i) / ∏_r (∑_y exp(−∑_j y_j b_j))^{n_r}
    for k in 1 : max_{r_i} i do
        L_k^{(r)} ← exp(−∑_j b_j r_i) / (∑_y exp(−∑_j y_j b_j))^{n_r}
    end for
    LR ← 2 · [L(y_ij, θ_i, b_j) − ∏_k L_k^{(r)}(y_ij, θ_i, b_j)]
    df ← (j − 1)(j − 2)
    p ← χ²_{(j−1)(j−2)}(LR)
    return [p, LR, df]
end procedure
```

and can be found in the package

To use the latter algorithm, the function can be exectued by the line

```
TidyLavaan::AndersenLRT(IRTmodel)
```

having an item response theoretic model as its input.

5.3 Polytomous Item Response Theory

A non-dichotomous, or **polytomous**, extension needs to be explored in order to fully capture the data without losing the ordinal structure. The Rasch model can be extended into a polytomous setting by means of the *Rating scale model*.

5.3.1 From rating scale model to the partial credit model

In the previous section, the works of Andersen proved to be a powerful method of checking the conditions of the dichotomous Rasch Model. Within the same year, Andersen wondered whether the idea of Rasch could be extended into a polytomous version (1973b). The idea of constructing a response vector y_{ij} is different in construction, as the vector denotes a partioning into each category, where $y_{ij}^{(c)}$ denotes whether the response y_{ij} is chosen for category m. For example, if participant $i = 43$ answer question $j = 6$ with $m = 5$, then $y_{43,6}^{(5)} = 1$ and $y_{43,6}^{(k))} = 0$ for $k \neq 5$. By Andersen 1973b, the polytomous Rasch model is then constructed by

$$\mathbb{P}(Y_{ij}^{(c)} = y_{ij}^{(c)} | \theta_i^{(c)}`, b_j^{(c)}) = \frac{\exp\left(y_{ij}^{(c)}(\theta_i^{(c)} - b_j^{(c)})\right)}{\sum_k \exp\left(\theta_i^{(k)} - b_j^{(k)}\right)} \quad \text{with} \quad m \neq k$$

By using the model displayed above, the problem of having categorical data is resolved. This idea has been extended into the **partial credit model**, in which the necessary assumption of having a uniform scale is relaxed. Therefore, it serves as a generalization of a polytomous Rasch model. The probability of attaining the next *step* within the likert scale is the foremost principle. To exemplify the notion of a step, consider the situation in which the score starts at 0. Then, the probability to reach the first step, or to take on the next score, 1, is defined by

$$\mathbb{P}(Y = m) = \frac{\pi_m}{\pi_{m-1} + \pi_m}$$

in which π_i is the unconditional probability of attaining the specific score m. In this case, the probability $\mathbb{P}(Y = m)$ is the conditional probability that the attained score is m given that $m - 1$ has already been attained. However, the expression of π_m is similar to the Rasch model, as

$$\mathbb{P}(Y = m) = \frac{\exp(\theta - b_m)}{1 + \exp(\theta - b_m)}.$$

Note the Markovian structure of b_m, as $b_{m-1} \leq b_m$ for $m = 1, \ldots, M$. The general formulation of π_m is given by

$$\pi_m = \frac{\exp(\sum_{k=0}^{m}(\theta - b_k))}{\sum_{l=0}^{M}\exp(\sum k = 0(\theta - b_k))}$$

and hence, if $\pi_{ij}^{(m)}$ denotes the probability that participants i has a realisation of the score m on a specific item j, then

$$\pi_{ij}^{(m)} = \frac{\exp(\sum_{k=0}^{m}(\theta_i - b_j^{(k)}))}{\sum_{l=0}^{M}\exp(\sum_{k=0}^{l}(\theta_i - b_j^{(m)}))}$$

5.4 Dimensionality

For every latent construct, the appropriateness of the inclusion of each item measuring the latent factor will be discussed in this section. Item response theory can asses the quality of each item loading onto the factor, similar to the structural equation model. Statistical appropriateness is not the only method of measurement to identify the proper questions. The **discriminality** parameter indicates how well a specific score on an item distinguishes a participant from other participants on a different position on the latent continuum of θ (ability). Mathematically, the parameter is described by the slope of the probability density with respect to the latent ability. If an item has little to no discriminability, It is best to discuss or analyse the possible lack of importance of inclusion in the model

5.4.1 Experience

For each latent construct, the results from the partial credit model will be given as well as additional analysis and suggestion to further improve the model ánd the survey

Experience t_1	**Discriminability parameter**
Gave me a feeling of beauty	1.844
Touched me	2.112
Appealed to my fantasy	0.701
Challenged by feeling for music	0.94
Provided inner rest	1.205
Made me experience a new part of the piece	0.541
Provided consolation	0.730
Made me feel sympathy	0.645
Made me lose track of time	0.840
Deeply impressed me	1.927
Made me able to clear my head	0.746

Analysis and suggestion The last column of the table compromises the discriminability parameter. The poorest items

- Appealed to my fantasy
- Made me experience a new part of the piece
- Provided consolation
- Made me able to clear my head
- Made me feel sympathy

did not provide satisfactory parameter estimates. On the other hand, the best items based on the discriminability parameter, ordered, are

1. Touched me
2. Deeply impressed me
3. Gave me a feeling of beauty

The implication of a high estimate does not imply a high score. It is merely a metric measuring the amount of information it can contain to distinguish a participant from a different participant. To exemplify, it was noted that *appealed to my fantasy* did not entail much discriminability into the model. By creating a simple histogram, this statistical measure seems to be intuitively true

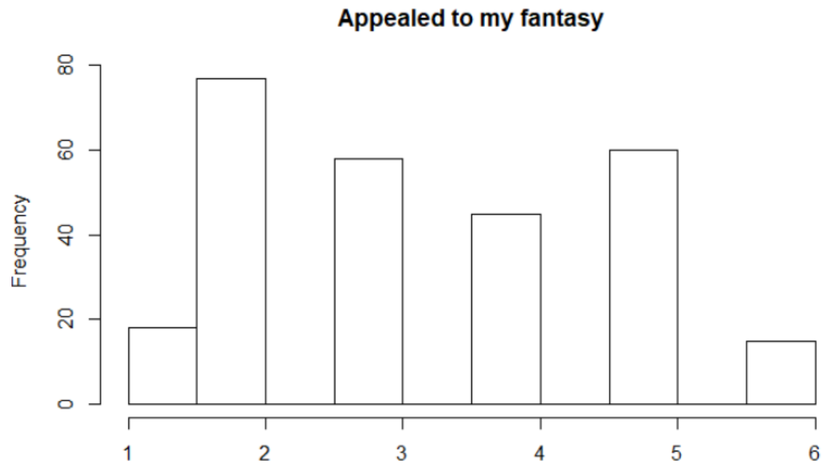

Figure 18: Distribution of *Appealed to my fantasy*

The data follows a uniform-esque distribution. On the other hand, it was stated that the item *touched me* provided the best discriminability parameter, which is in line with the histogram displayed below.

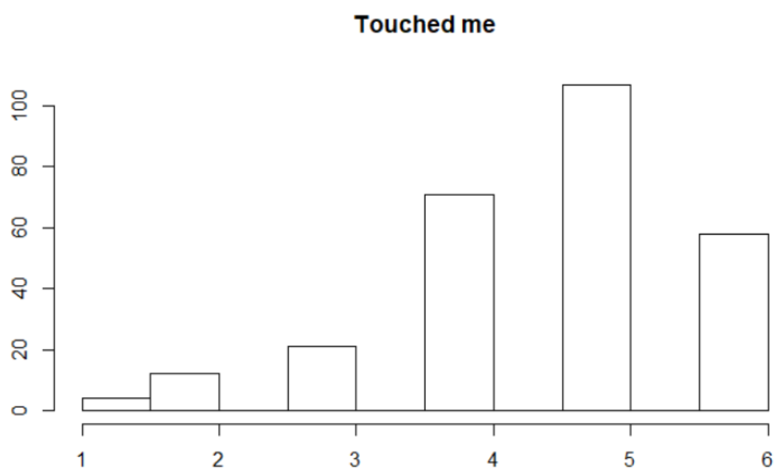

Figure 19: Distribution of *Touched me*

Indeed, the satisfying distribution intuitively confirms the results from the partial credit model. For the experience at time 2, a similar result is expected. By computing the same table,

Latent variable	**Discriminability**
Gave me a feeling of beauty	1.447
Touched me	1.862
Appealed to my fantasy	0.715
Challenged my feeling for music	1.215
Provided inner rest	1.902
Made me experience a new part of the piece	0.701
Provided consolation	0.894
Made me feel sympathy	0.786
Made me lose track of time	0.919
Deeply impressed me	1.478
Made me able to clear my head	0.984

5.4.2 Mood

Specifying the same parameter, the discriminability parameters of mood tell a different story. First, consider the mood at time 1. The data shows

Item	Discriminability
Difficult	0.333
Surprising	0.686
Relaxing	0.626
Emotional	1.151
Exciting	0.709
Demanding	0.003
Beautiful	3.455

Immediately, two items are of interest. First of all, the estimate for *Demanding* is really poor and most definitely is not significantly different from zero. On the other hand, the item *beautiful* scores extremely well. After experience the *Matthaus Passion* from Bach in person, it is no surprise that the music of Bach is easily found to be both emotional and beautiful at the same time. The results are quite similar for the mood at the second musical piece, as

Item	Discriminability
Difficult	-0.144
Surprising	1.319
Relaxing	1.908
Emotional	1.511
Exciting	1.920
Demanding	0.23
Beautiful	3.499

5.4.3 Meaning and Interpretation

As the meaning and interpretation has proven to be an important factor in the model, it is noteworthy to analyse the results from the item response theory model. At time 1, it can be obtained that

Item	Discriminability
I recognized the emotions in the music	2.209
I felt appealed by the developments in the musical piece	2.785
I was able to experience the intentions of the composer	1.551
I tended to physically move	0.596
I listened to the music primarily with my eyes closed	0.404

Similarly, for time 2,

Item	Discriminability
I recognized the emotions in the music	2.552
I felt appealed by the developments in the musical piece	2.520
I was able to experience the intentions of the composer	1.937
I tended to physically move	0.530
I listened to the music primarily with my eyes closed	0.344

Again, great similarity between the two tables can be noticed. The discriminability parameter estimates tell a diverse story, as the items

- *I tended to physically move*; and
- *I listened to the music primarily with my eyes closed*

indicate a poor discriminability. As the factor loadings in the structural equation model were poor (although not significantly poor), the item response theory creates an extra layer of evidence against using such an item.

5.5 The inertia to agree with an item

As listed in the mathematical models, not only the discriminability parameter is estimated. The *difficulty*-parameter is a central topic of the item response theory. Recall that the difficulty paramater can be interpreted as *the hardness to agree with a question or item*.

By means of extracting the estimated difficulty parameter found extracted by functions in the eRm package, the standard error and confidence intervals for the estimates can be calculated. Within the TidyLavaan package, this process can simple be computed by

```
TidyLavaan::ConfidenceIntervalIRT(IRTmodel)
```

Both a rating scale model and a partial credit model are admissible inputs for this function. The output is a data frame, with the category, the estimated difficulty, the standard error (denoted by se) and both the lower and upper bound of the confidence interval.

First, consider the mood at time 1 and time 2 respectively

Category	Difficulty	SE	2.5%	97.5%
Happiness	-0.52	0.06	-0.64	-0.39
Energetic	0.06	0.05	-0.04	0.17
Excited	0.86	0.06	0.74	0.98
Enthusiastic	-0.44	0.06	-0.56	-0.32
Emotional	1.2	0.07	1.06	1.33

Table 8: Estimation of the difficulty parameter at time 1

Category	Difficulty	SE	2.5%	97.5%
Happiness	-0.77	0.08	-0.93	-0.62
Energetic	-0.36	0.07	-0.5	-0.22
Excited	1.54	0.09	1.37	1.71
Enthusiastic	-0.14	0.07	-0.27	0
Emotional	1.46	0.09	1.29	1.63

Table 9: Estimation of the difficulty parameter at time 2

As the numerical interpretation does not follow trivially from the table, the scale and numerical assignment of the estimator difficulty parameter b_j should be discussed first. As stated before, the latent b_j represents the difficulty of a particular item. For negative values of b_j, the item relatively easy to agree with. For a value of b_j close to zero, the items have a neutral difficulty. For b_j greater than zero, the question was relatively hard to agree with. The author would like to emphasize the intuitive meaning of *difficulty to agree with an item* rather than the plain difficulty of an item. Then, it can be obtained that it was more easy to score higher on the items *happiness, energetic, emotional*, but it was easier to score less high on the items *excited, enthusiastic and emotional*. It is only reasonable to expect changes over time, but by using the confidence interval, the results show that the all items have significantly changed over time.

Within the concert, the experience of a classical piece at each specific time has been measured by means of the items related to meaning and interpretation as described in the structural model. At both the distinct classical pieces, questions on the meaning and interpretation can yield insights in the change of experience. Conducting a similar study and by using the items which were discussed in the discriminability section, we have

Category	Difficulty	SE	2.5%	97.5%
I felt addressed by the developments in the musical piece	-0.54	0.06	-0.66	-0.42
I experienced the meaning as the composer had intended	0.12	0.05	0.02	0.22
I tended to physically move during the classical piece	0.5	0.05	0.4	0.61
I listened to the music primarily with my eyes closed	0.71	0.06	0.6	0.82

And at time 2

Category	Difficulty	SE	2.5%	97.5%
I felt adressed by the developments in the musical piece	-0.58	0.06	-0.70	-0.46
I experienced the meaning as the composer had intended	0.05	0.05	-0.05	0.12
I tended to physically move during the classical piece	0.56	0.05	0.45	0.67
I listened to the music primarily with my eyes closed	0.62	0.06	0.51	0.73

In contrast to the previous results, all items did not significantly change over time. For both musical pieces, it was relatively easy to feel addressed by the developments in the musical piece. However, people felt indifferent about the identification with the composer. Even more, the participants were not easily physically moved nor listened to the music with their eyes closed. This is fully similar to the results which were found in the structural equation model.

Finally, the items measuring experience should be measured and analysed. Again, constructing tables with parameter estimates yields

Category	Difficulty	SE	2.5%	97.5%
Moved me	-0.67	0.07	-0.8	-0.54
Appealed to my fantasy	0.62	0.06	0.51	0.74
Challenged my feelings of music	-0.1	0.06	-0.22	0.01
Created inner peace	-0.33	0.06	-0.46	-0.21
Made me aware of something I did not know	0.29	0.06	0.17	0.4
Provided Consolation	0.87	0.06	0.75	0.99
Provided sympathy	0.91	0.06	0.79	1.03
Made me lose track of time	-0.22	0.06	-0.34	-0.1
Deeply impressed me	-0.33	0.06	-0.45	-0.21
Made me able to clear my head"	-0.16	0.06	-0.27	-0.04

Also, for the second classical music piece,

Category	Difficulty	SE	2.5%	97.5%
Moved me	-0.77	0.07	-0.91	-0.62
Appealed to my fantasy	0.8	0.06	0.67	0.92
Challenged by feelings of music	-0.07	0.06	-0.19	0.06
Created inner peace	-0.39	0.07	-0.52	-0.25
Made me aware of something I did not know	0.29	0.06	0.17	0.41
Provided consolation	0.87	0.06	0.74	0.99
Provided sympathy	1.07	0.07	0.94	1.2
Made me lose track of time	-0.17	0.06	-0.3	-0.05
Deeply impressed me	-0.45	0.07	-0.58	-0.31
Made me able to clear my head	-0.14	0.06	-0.27	-0.02

Then, it can be observed that the itmes

- *Moved me*;
- *Created inner peace*;
- *Made me lose track of time*;

- *Deeply impressed me*; and
- *Made me able to clear my head*

were relatively easy to agree with. On the other hand, the items

- *Appealed to my fantasy*;
- *Made me able to clear my head*; and
- *Provided sympathy*

were relatively hard to agree with. Once more, the results are in line with the structural equation modeling providing evidence for having constructed a solid model.

5.6 Re-integration with the structural model

The structural equation model remains the main method for this thesis. Item response theory is solely a mean of improving the model. Some factors and item shine in the structural model, some are poor, and some are in a grey area. For the latter category, an alternative measure provides a more diverse insight. As important notions on micro-level have been established this chapter, it is time to re-integrate the results into a statistically improved model. The following items were removed as they did not indicate a good fit within **both** the structural model and the item response model

1. I primarily listened to the music with my eyes closed
2. In a hurry
3. Demanding
4. I tended to physically move
5. I experienced something new in the piece which I did not know
6. Made me able to clear my head

For all of these items, at both times the fit has been established as a poor fit. The fit indices of the structural equation vastly improved, as the new fit incides show

- CFI : 0.96 (+)
- TLI : 0.96 (+)
- RMSEA : 0.048 (+)

vastly improving the model into an excellent fit. Now, the regressions shows

Latent Variable	**Correlates with**	**Standard Error**	**Estimate**
Meaning and Interpretation t_1	Knowledge	0.132	-1.409
Meaning and Interpretation t_2	Meaning and Interpretation t_1	0.091	0.104*
Meaning and Interpretation t_2	Knowledge t_1	0.142	1.012*
Mood t_2	Mood t_1	0.036	0.838
Experience t_1	Meaning and Interpretation t_1	0.234	0.281*
Experience t_1	Mood t_1	1.593	-3.386
Experience t_2	Meaning and Interpretation t_2	0.175	1.367
Experience t_2	Mood t_2	0.350	1.257

Table 10: Results of latent factor score improvement per partition of knowledge

showing significant changes in the regression coefficients. Similar interpretations can be given, however. As visualised in the previous chapter, the influence knowledge exerts on the meaning and interpretation is negative. Also, the standard error is relatively small indicating a valid regression coefficient. Similarly, the regression coefficient for mood at experience at time 2 is in line with the results established thus far. Hence, for every increase of score in the mood, the score of experience increases with an eight of a point. Nevertheless, it should be noted that more regressions are poor due to having a high standard error. Therefore, some interdependencies in the model cannot be successfully established. Most importantly, the regression of experience at time 1 and mood at time give troublesome results due to having numerical difficulty.

5.6.1 Proper scientific conduct

The implications of the model and the survey are as follows. It is only reasonable to criticize the purely statistical basis to improve the model. Of course, for every survey is it possible to find item indicating a poor fit. Still, these question might give relevant information on its own without consideration for a structural model. Therefore, it is not recommended to fully drop variables or to reason purely with statistical significance. Also, the exclusion of a variable is prone to a **type-I** error, as exclusion is funded upon the p-value of the factor loading (and whether is significantly different from zero, usually with $\alpha = .05$). Thus far, structural equation models and item response theories have been discussed. Whereas a structural model discusses the interdependencies of the latent constructs, the item response theory only considers one factor at a time. Therefore, the results of both method cannot be fully similar. Nevertheless, a great similarity of the two methods is expected. As item response theory also compromises a method of factor estimation, a great insight is the correlation between the respective factor scores. By a simple computation, it can be obtained that

Latent variable	**Correlation**
Experience t_1	0.946
Experience t_2	0.948
Meaning and Interpretation t_1	0.960
Meaning and Interpretation t_2	0.955
Mood t_1	0.911
Mood t_2	0.896
Knowledge	0.920

Which is exactly to be expected. This results provides confidence in the result of the study, as different methods present similar results, but from a different perspective.

6 Discussion

Mathematical analysis can sometimes be a lot to take in. This section tries to recap and critically discuss each chapter for each result. At the end, the reader should be able to understand the chosen methodology, the interpretation of the results and the suggestions for conduction a similar survey.

6.1 Methodology

Although no statistical method is without its critics, the structural equation model provided the right method for determining dimensionality, the quality of measurement and the interdependencies of latent variables, tackling the problems posed as the aim for this study. In order to ensure the model is properly identified, two important problems arising from a structural equation model have been discussed. First, a switch to the weighted least squares estimator has been made to account for the non-normality the data omits. Secondly, by means of making use of the modification index, the model specification was studied and improved. Although these two possible problems were not the main interest of this thesis, they should be discussed to ensure a thorough study.

6.2 Dimension reduction

For about half of the data, a very good method for data reduction has been found. More specifically, 62 items were reduced into seven factors. Fortunately, the items and factors within this have intuitively seem to be the more important variables. More conceretely, the following result was found

Factor	Time	Inclusion
Reasons of visiting the theatre	0	No.
Knowledge about the composers	0	Yes
Setting of the concert	all	No. Meausured well, but no good addition to the model
Meaning and Interpretation	1,2	Yes
Mood	1,2	Yes
Experience	1,2	Yes
Break	1,2	No. Not well-defined nor numerically valid to include
Musical background	all	No. Numerical Problems
Quality of the musicians	all	No. Little to no variability

It should be noted that not being able to include a factor into the structural model does not necessarily imply a poorly measured factor. For instance, the quality of the musicians was too highly scored such that the variability among its items was too low to properly include in the model. On the other hand, the musical background of the participants was a binary-evaluated variables, which does not pose a well founded inclusion in the model but could help to analyse the data further. In order to construct a more nuanced view, the items were studied by both a structural equation model and and item response model. The six items

1. I primarily listened to the music with my eyes closed
2. In a hurry
3. Demanding
4. I tended to physically move
5. I experienced something new in the piece which I did not know
6. Made me able to clear my head

were excluded after the integration of the two methods as they lacked an adequate fit in both models.

6.3 Quantifying and assessing factor structures

Thus far, this study mainly discussed the results of the items having a good quality. However, as shown in the previous table, it is also possible to identify the problems with the poorly fitted factors. Two interesting results from the regression should be listed.

1. *Knowledge* exerts a positive influence upon the experience.
2. *Knowledge* exerts a negative influence upon the mood and interpretation. The interpretation and analyses of this result is an example of conducting a psychological- or sociological study.

This conclusion is supported both inspecting the factor score estimations and the regression coefficients. Nevertheless, it should be noted that the regression coefficient of mood at time 1 is worrying and not in line with the factor score estimation procedures. This problem certainly needs to be addressed when wanting to make use of this specific regression coefficient.

6.4 Development of a generalizable R-package

As the methods used in this study are usually applied by the social sciences, most statistical software in R can be quite intimidating. Therefore, a parallel aim of this study was to develop a R-package constructing a structural equation model easier. More precisely, the following difficulties were solved

1. An ETL-proceds for handling survey data, which comprises:
   - Seeks possible numerically difficulties such as having fully linearly dependent columns, resulting in a non positive definite matrix
   - Structurizing the data by means of meta data, consisting of each factor with its respective factor.
   - Handling ordinal data
2. Functions providing great usability for implementing a Lavaan model, comprising
   - Measurement model.
   - Measurement model with restricted factor loadings.
   - Creating the $\boldsymbol{\Psi}$ matrix in order to correlate items.
   - An alternative factor estimation process
   - Functions detecting a change over time in factor score estimation procedures
3. Functions providing great usability for implementing an item response theory
   - Automatically generate item response theory for every factor present in the data
   - Functions detecting change over time in factor score estimation procedures

6.5 Improvements and acknowledgements

A few critical acknowledgements should be made. First of all, the data at hand did require a manual ETL step - resulting in a deviation from the raw source of the data. Secondly, due to numerical problems, the Wald test for the structural equation model could not be assessed. Although the modification index yields possible inclusions for the model, possible exclusions were not able to be tested due to numerical difficulty.

At time of publishing this thesis, the author is unable to publish the package yet. However, at the time of writing (July 2020), the author expects the package to available through GitHub and possibly CRAN.

References

[1] D Kaplan (2008), Structural equation modeling: Foundations and Extentions, *Sage*,

[2] Hui,C., Law, K.S. and Chen Z.X. (1999), A structural equation model of the effects of negative affectivity, leader-member exchange, and perceived job mobility on in-role and extra-role performance, *Elsevier*, p 3-21

[3] Meyer, J.P. and Zhu, S. (2013), Fair and equitable measurement of student learning in MOOCs: An introduction to item response theory, scale linking, and score equating, *ERIC*, p28-36

[4] Wright (1921), Correlation and causation, *J. Agricultural Research.* 557–585.

[5] Lo, D.W., Stewart, W.P. (2002), A Structural Equation Model of Residents' Attitudes for Tourism Development

[6] Tempelaar, Van der Loeff and Gijselaers (2007), A structural equation model analyzing the relationship of students' attitudes toward statistics, prior reasoning abilities, and course performance. *Contemporary Educational Psychology 32*, p 105-131

[7] Andersen, BE (1973), conditional inference for multiple-choice questionnaires. *The British Psychological Society*

[8] Musil, Jones, S.L. and Warner, C.D. (1998), Structural equation modeling and its relationship to multiple regression and factor analysis. *Research in nursing and Health 21*, p 271-281.

[9] Richter, N., Sinkovics, R.R., Ringle, CM, and Schlägel, (2016). A critical look at the use of SEM in international business research, *International Marketing Review*, 33(3), p376-404.

[10] Fan,J., Ke, Z.T., Liu, H. and Xia, L (2014), A supervised dimension reduction method via Rayleigh Quotient Optimazation, *the analysis of statistic 43 (4)*, p 1503-1509

[11] Boomsma, A. (1999), Book review of K.A. Bollen (1989). Structural Equations with latent variables. *Kwantitative methoden 12 (38)*, p 124-131.

[12] Bollen, K.A. (1989) Structural equations with latent variables, *Wiley*

[13] Glass,G. and Maguire, T. Abuses of Factor scores (1996), *American Education Research journal*, p 297-303

[14] T Asparouhova, T. and Muthen, B. Structural Equation Modeling: A Multidisciplinary Journal (2014), p 1-4

[15] Azevedo,K. Using factor score estimates in latent variable analysis (2002), Retrospective Theses and Dissertations. p1-149

[16] Petrillo J, Cano SJ, McLeodLD, Coon C (2015). Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: a comparison of worked examples, Value Health

[17] Baker, J.G., Rounds, J.B. and Zevon M.A. (2000) A Comparison of Graded Response and Rasch Partial Credit Models with Subjective Well-Being, *Journal of Educational and Behavioral Statistics 25 (3)*, p 253-270

[18] Hatzinger, R. (2008). A GLM framework for item response theory models. (2008), *Reissue of 1994 Habilitation thesis*, p 13-46

[19] Schumm, W.R.,Pratt, K.K. and Hartenstein, J.L. (2013). Determining statistical significance (alpha) and reporting statistical trends: Controversies, issues, and facts. *Comprehensive psychology*

[20] Glöckner-Rist, A., and Hoijtink, H. (2003). The best of both worlds: Factor analysis of dichotomous data using item response theory and structural equation modeling. *Structural Equation Modeling, 10(4)*, 544–565.

[21] Raju, N. S., Laffitte, L. J., and Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. Journal of Applied Psychology, 87(3), 517–529.

[22] Sang Gyu Kwak Jong Hae Kim, 2017, Central limit theorem: The cornerstone of modern statistics. *Korean journal of anesthesiology 70(2):144*

[23] Hayton, D Allen, V Scarpello, 2010, Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis, *Organizational Research Methods 7(2):191-205*

West, S. G., Finch, J. F., Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications (p. 56–75). Sage Publications, Inc.*

[24] Boomsma 1983, Nonconvergence, improper solutions, and starting values in lisrel maximum likelihood estimation, *PhD thesis*

[25] Amrhein, Greenland and McShane, 2019, cientists rise up against statistical significance, *nature*

[26] Silvey, S. D. The Lagrangian Multiplier Test. *Ann. Math. Statist. 30 (1959), no. 2, 389–407.*

[27] Gorassini, D. R., Spanos, N. P. (1986). A social-cognitive skills approach to the successful modification of hypnotic susceptibility. *Journal of Personality and Social Psychology, 50(5), 1004–1012*

[28] Hu, L.-T., Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, issues, and applications (p. 76–99). *Sage Publications, Inc.*

[29] Bagozzi, Yi, 1991, Assessing Construct Validity in Organization Research, *Administrative Science Quarterly 36:421-458*

[30] Ten Berge, Krijnen, 1999, Some new results on correlation-preserving factor scores prediction methods, *Linear Algebra and its Applications Volume 289, Issues 1–3, 1 March 1999, Pages 311-318*

[31] Herschberger, 2005, Factor Score Estimation, *Encyclopedia of Statistics in Behavioral Science*

A Proofs

A.1 Proof of the Rayliegh coefficient

Before using the Rayleigh quotient, consider first the decomposition of x in terms of a basis of the eigenvectors v_i

$$x = \boldsymbol{a} \cdot \boldsymbol{v} = \sum_{i=1}^{n} a_i v_i$$

where

$$a_i = \frac{\langle x, v_i \rangle}{\|v_i\|^2}$$

in order to establish orthogonality. Then, by using the established LU decomposition,

$$\begin{aligned} R(\boldsymbol{\Sigma}, x) &= \frac{x^T L^T L x}{\|x\|^2} \\ &= \frac{\left(\sum_{i=1}^{n} a_j v_j\right)\left(\sum_{i=1}^{n} a_i \lambda_i v_i\right)}{\sum_{i=1}^{n} a_i^2 \|v_i\|^2} \end{aligned}$$

By orthonormality of v_i, the latter reduces to

$$= \sum_{i=1}^{n} \lambda_i \frac{(x^T v_i)^2}{\|x\|^2 \|v_i\|^2}$$

which attains its maximum whenever $v \in E$ where E is the eigenspace corresponding to $\lambda_{\max} = \max \operatorname{spectrum}(\boldsymbol{\Sigma})$.

A.2 Proof of the identities in the ML-derivation

Proof:

1. By plain computation

$$\operatorname{trace} AB = \sum_{i=1}^{n} (AB)_{ii} = \sum_{i=1}^{n} \sum_{k=1}^{m} A_{ij} B_{ji} = \sum_{k=1}^{m} (BA)_{jj} = \operatorname{trace} BA \tag{10}$$

2. as $x^T A x$ is a scalar, it is invariant under taking its trace

3. Again, by computation

$$\operatorname{trace} AB = \sum_{i=1}^{n} a_{1i} b_{i1} + \sum_{i=1}^{n} a_{2i} b_{i2} + \cdots + \sum_{i=1}^{n} a_{mi} b_{im}$$

Implying

$$\frac{\partial}{\partial A} \sum_{i=1}^{n} a_{1i} b_{i1} + \sum_{i=1}^{n} a_{2i} b_{i2} + \cdots + \sum_{i=1}^{n} a_{mi} b_{im} = b_{ji} = B^T$$

B Code

B.1 ETL

For the ETL-proces (Extract, transform and load), the following functions were constructed and used at different stages of the study

```
### Data inladen
library(readxl)
library(dplyr)
library(lavaan)
library(Matrix)
library(plm)
library(caret)
library(matrixcalc)
library(scales)
library(semPlot)
library(FactoMineR)
library(psych)
library(lavaanPlot)

Clean_Colnames <- function(data, RepeatedMeasures){
  # as special characters might result in problems in the program,
  # they will be removed
  s <- colnames(data) %>%
    gsub('[[:digit:]]+', '',.) %>%
    gsub( "[^[:alnum:]///' ]", "", .) %>%
    gsub('/', "",.) %>%
    gsub(' ', "",.)
  if (RepeatedMeasures == T){
    # If column names are not unique, they will be enumerated (x, x1, x2, ..)
    s <- s %>%
      make.unique(., sep = '_')
  }
  return(s)
}

# automatically creates a meta data table, an improved summary
# specific for survey data
MetaData <- function(data){
  Classes <- data %>% sapply(., class) %>% as.data.frame()
  scale0 <- data %>%
    sapply(., min) %>% as.data.frame() %>% pull()
  scale1 <- data %>%
    sapply(., max) %>% as.data.frame() %>% pull()
  mean1 <- data %>%
    sapply(., mean) %>% as.data.frame() %>% pull()
  return(data.frame(Column = colnames(mydata), Minimum = scale0, Maximum = scale1, Mean =
}

# Converting to numeric might result in NA's. This tries to overcome that
ParseNumeric <- function(x) {
  return(suppressWarnings(sum(is.na(as.numeric(x)))) == sum(is.na(x)))
}

# Whenever the previous function returns true, the variable will be converted to numeric
Convert <- function(data){
  return(data %>%
```

```
            mutate_if(ParseNumeric, as.numeric))
}

# Questions should be reverse coded sometimes
Reverse_Question <- function(column){
  column <- -1*column
  mini <- min(column)
  column<-column + abs(mini)
  return(column)
}
```

B.2 Structural equation model

```
mydata <- mydata %>%
  mutate_at(c('moeilijk', 'moeilijk_1', 'Ikwasmaarnetoptijdindegrotezaalaanwezigvoordeaan
            'Erwarenstorendefoutenindeuitvoeringvandemuziekstukken', 'Ikwasmetijdenshe
            ), ~Reverse_Question(.))

formule_bouwen <- function(factors, ConstraintLogical, ConstraintLabel){
  y <- unique(factors[,1])
  factor <- pull(factors[,2])
  paste(y, "=~", paste(factor, collapse=" + "))
}

MeasurementModel <- function(MetaData, SelectedFactors){

  meta_list <- split(MetaData, MetaData$Factor)[SelectedFactors]

  model <- meta_list %>%
    lapply(., dplyr::select,Factor, Item) %>%
    lapply(., formule_bouwen, ConstraintLogical = F) %>%
    paste(., collapse = '\n\n')

  return(model)
}

formule_bouwen <- function(factors, ConstraintLogical, ConstraintLabel){
  y <- unique(factors[,1])
  factor <- pull(factors[,2])

  if (ConstraintLogical == TRUE){
  CS <- paste0(ConstraintLabel, 1:length(factor))
   output <- paste(y, "=~", paste(paste0(CS, '*', factor), collapse=" + "))
  } else{
   output <- paste(y, "=~", paste(factor, collapse=" + "))
  }
  return(output)
}

MeasurementModelConstraint <- function(MetaData, SelectedFactors, ConstraintLogical, Con
  meta_list <- split(MetaData, MetaData$Factor)[SelectedFactors]

  model <- meta_list %>%
    lapply(., dplyr::select,Factor, Item) %>%
    lapply(., formule_bouwen, ConstraintLabel = ConstraintLabel, ConstraintLogical = Con
    paste(., collapse = '\n\n')
```

```
  return(model)
}

VarianceModel <- function(MetaData,x,y){

  meta_list <- split(MetaData, MetaData$Factor)[c(x,y)]
  output <- paste(meta_list[[x]]$Item, "~~", meta_list[[y]]$Item, collapse = '\n') # pas
  return(output)
}

FirstModel <- paste0(
  MeasurementModelConstraint(meta_data, c('BetekenisenInterpretatietijdstip1', 'Betekeni
  MeasurementModelConstraint(meta_data, c('GemoedstoestandTijdstip1', 'GemoedstoestandTi
  MeasurementModelConstraint(meta_data, c("Ervaringtijdstip1", "Ervaringtijdstip2"), Con
  MeasurementModel(meta_data, c("Kennis")), '\n\n',
  VarianceModel(meta_data, "BetekenisenInterpretatietijdstip1", "BetekenisenInterpretati
  VarianceModel(meta_data, "Ervaringtijdstip1", "Ervaringtijdstip2"), '\n',
  VarianceModel(meta_data, "GemoedstoestandTijdstip1", "GemoedstoestandTijdstip2"), '\n\


  'BetekenisenInterpretatietijdstip1 ~ Kennis \n\n',
  'BetekenisenInterpretatieTijdstip2 ~ BetekenisenInterpretatietijdstip1\n\n',
 'GemoedstoestandTijdstip1 ~ GemoedstoestandTijdstip2 \n\n' ,
  'Ervaringtijdstip1 ~ GemoedstoestandTijdstip1 + BetekenisenInterpretatietijdstip1
\n\n',
 'Ervaringtijdstip2 ~  GemoedstoestandTijdstip2 + BetekenisenInterpretatieTijdstip2', '\
  '\n\n', adjusted_model
)


summary(FirstLavaan)

cat(FirstModel)
FirstLavaan <- sem(FirstModel, mydata)
FirstLavaan
fitmeasures(FirstLavaan)
summary(FirstLavaan)

add_cov <- modificationindices(FirstLavaan) %>%
  filter( op == '~~') %>%
  filter(mi > 10)


adjusted_model <-  paste(paste(as.vector(paste(add_cov$lhs, add_cov$op, add_cov$rhs)), "

data.frame(FirstLavaan@ParTable$lhs, FirstLavaan@ParTable$rhs, FirstLavaan@ParTable$est)
  as_tibble %>%
  `colnames<-`(c('Latent Variable', 'Item', 'Coefficient')) %>%
  mutate(Coefficient = round(Coefficient, digits = 3)) %>%
  write.csv2(., 'RegressionOutput.csv')


  LavPredictKrijnen <- function(LavaanModel){

  library(psych)
```

```
  MatrixSquareRoot <- function(x) {
    e <- eigen(x)
    e$values[e$values < 0] <- 0
    return(e$vectors %*% diag(sqrt(e$values)) %*% t(e$vectors))
    }


  InverseMatrixSquareRoot <- function(x) {
    e <- eigen(x)
    e$values[e$values < 0.00001] <- 0.00001
    return(e %>% `[[`('vectors') %*% diag(1/sqrt(e %>% `[[`('values'))) %*% t(e %>% `[[`
    }

  Data <- inspect(LavaanModel, what="data")
  Lambda <- inspect(LavaanModel, what="std")$lambda
  Phi  <- inspect(LavaanModel, what = 'std')$psi %>% as.data.frame() %>% as.matrix()

  if(missing(Phi)) {
    Phi <- diag(1, ncol(Lambda))
  } else {
    Phi <- Phi
  }

  L <- Lambda %*% MatrixSquareRoot(Phi)

  rMatrix <- Data %>%
    cor(., use = 'pairwise') %>%
    MatrixSquareRoot() %>%
    solve()


  if (corpcor::is.positive.definite(cor(Data), tol = 10^{-6}, method = 'eigen') == TRUE)
    rInverse <- cor.smooth(cor(Data)) %>%
      solve()
  } else{
    rInverse <- cor.smooth(cor(Data)) %>%
      solve()
  }

  C_d <- rMatrix %*% L %*% InverseMatrixSquareRoot(t(L) %*% rInverse %*% L)
  W <- rMatrix %*% C_d %*% MatrixSquareRoot(Phi)


  Weights <- W %>%
    `colnames<-`(colnames(Lambda)) %>%
    `rownames<-`(rownames(Lambda))

  FactorScores <- scale(Data) %*% Weights
  return(FactorScores %>% as_tibble)

}

DetectChange <- function(Kolom1, Kolom2){
  StandardError <- sd(Kolom1)/sqrt(length(Kolom1))
  LowerBound <- Kolom1 - 2.576 * StandardError
```

```
  UpperBound <- Kolom1 + 2.576 * StandardError
  SignificantChange <- ifelse(Kolom2 >= LowerBound & Kolom2 <= UpperBound, 0, 1)

  output <- data.frame(
   Time1 = Kolom1, Time2 = Kolom2,
   `Lower Bound` = LowerBound,
   `Upper Bound` = UpperBound,
   `Significant Change` = SignificantChange
  )
  return(output)
}
```

B.3 Item respones theory

```
library(ltm)
library(mirt)
library(dplyr)
library(rlist)
library(readxl)
library(ggplot2)

datatfunc <- function(data){
  if (data %>% sapply(., max)  %>% max() > 8){
    data1 <- data %>%
      sapply(., rescalefun, 0,6) %>% as.data.frame() %>%
      sapply(.,round)
  } else{
    data1 <- data
  }
  return(as_tibble(data1))
}

t <- function(x){return(ncol(x) > 4)}

listfunc <- function(data){
  return(data %>% sapply(., function(x){return(ifelse(x < 4, 0, 1))}))
}


FilterNrow <- function(datalist, minimum_rows){
  elements <- which((datalist %>% lapply(., ncol) %>% as.data.frame() > minimum_rows))
  return(datalist[elements])
}

unique_filter <- function(x){
  if (x %>%  sapply(., unique) %>% class() == "matrix"){
    return(x)
  } else if (x %>%  sapply(., unique) %>% class() == "list"){
    fil <- x %>%
      sapply(., unique) %>%
      sapply(., length)
    return(x[fil > 2])
  } else{
    warning("Possibly invalid data, please check results carefully")
  }
}
```

```
Extract_Difficulty <- function(Model){
  diff <- Model$betapar %>% as.data.frame()
  DifficultyDF <- data.frame(Item = rownames(diff), Difficutly = diff)
  rownames(DifficultyDF) <- NULL
  colnames(DifficultyDF) <- c("Item", "DifficultyParameter")
  return(DifficultyDF %>% as_tibble())
}

rescalefun <- function(x,a,b){
  return(scales::rescale(x, to = c(a,b)))
}

# Apply meta list to the data
    DataListPerCategory2 <- NULL
meta_list2 <- split(meta_data2, meta_data2$Factor)
for (i in 1:length(meta_list2)){
  DataListPerCategory2[[i]] <- mydata[, meta_list2[[i]]$Item]
}
names(DataListPerCategory2) <- names(meta_list2)

DataListPerCategory2 <- DataListPerCategory2[c('Ervaringtijdstip1', 'Ervaringtijdstip2',
                                               'BetekenisenInterpretatietijdstip1', 'Bete
                                               'GemoedstoestandTijdstip1', 'Gemoedstoesta
                                               'Kennis')]
PCMmodel <- DataListPerCategory2 %>%
  lapply(., datatfunc) %>%
  lapply(., unique_filter) %>%
  FilterNrow(., 3) %>%
  lapply(., as_tibble) %>%
  Filter(nrow,.) %>%
  lapply(.,eRm::PCM)

  ConfidenceIntervalIRT <- function(Model){
  Eta <- Model$etapar %>% as.data.frame()
  Etadf <- data.frame(
   Categorie = rownames(Eta),
   Difficulty = Eta,
   SE = Model$se.eta
  )
  rownames(Etadf) <- NULL
  colnames(Etadf) <- c("Categorie", "Difficulty", "SE")

  Final_df <- Etadf %>%
    mutate(`2.5%-bound` = Difficulty + stats::qnorm(0.025)*SE) %>%
    mutate(`97.5%-bound` = Difficulty + stats::qnorm(0.975)*SE)
  return(Final_df)
}

CIRSM <- PCMmodel %>%
  lapply(., ConfidenceIntervalIRT) %>%
  lapply(., filter, !grepl('Cat', Categorie))

Afname <- CIRSM$Ervaringtijdstip2 %>%
  mutate(Difficulty = round(Difficulty, digits = 2)) %>%
  mutate(SE = round(SE, digits = 2)) %>%
  mutate(`2.5%-bound` = round(`2.5%-bound`, digits = 2)) %>%
```

```
  mutate(`97.5%-bound` = round(`97.5%-bound`, digits = 2))

Ability <- data.frame(
  Persoon = round(rowSums(dat[9:13])/5),
  Tijdstip1 = fscores(T1, full.scores.SE = TRUE) %>% as.data.frame(),
  Tijdstip2 = fscores(T2, full.scores.SE = TRUE) %>% as.data.frame()
) %>%
  `colnames<-`(c("PredefinedKnowledge", "Tijdstip1", "SE1", "Tijdstip2", "SE2")) %>%
  mutate(Verschil = Tijdstip1 -Tijdstip2) %>%
  mutate(GecombSE = sqrt(SE1^2 + SE2^2)/sqrt(2))

resultdf <- data.frame(
  Mean1 <- mean(Ability$Tijdstip1),
  Mean2 <- mean(Ability$Tijdstip2),
  SE12  <- sqrt(sum(Ability$GecombSE^2)/sqrt(289))
) %>%
  `colnames<-`(c("Mean1", "Mean2", "SE")) %>%
  mutate(verschil = Mean1 - Mean2)

  AndersenLRT <- function(IRTmodel){


fulllog <- IRTmodel$loglik

splittedmodel <- IRTmodel$X %>%
  as.data.frame() %>%
  mutate(Totals = rowSums(.)) %>%
  mutate(Totals = round(round(round(Totals)/10)/.1, digits = 1)) %>%
  split(., .$Totals) %>%
  Filter(function(x){nrow(x) > 5},.) %>%
  lapply(., dplyr::select, -Totals) %>%
  lapply(., PCM)

extractloglikelihood <- function(model){
  return(model$loglik)
}

partitionedlog <- splittedmodel %>%
  lapply(., extractloglikelihood) %>%
  do.call('rbind',.) %>%
  sum()

AndersenLRT(PCMmodel)

PCMmodel[[2]]$log.Lik


LR =  2* (partitionedlog - fulllog)

J <- length(splittedmodel) + 1
p_value <- 1 - pchisq(LR, (J-1) * (J-2))

OutputList <- list(fulllog, partitionedlog, LR,  (J-1)*(J-2), p_value)
names(OutputList) <- c("Loglikelihood full model", "Loglikelihood partitioned model", "A
return(OutputList)
}
```