Applying Structural Equation Modelling on a Motivation Survey

by

Linda Leeuwestein

Student Name:	Student Number:
Linda Leeuwestein	4467744

Instructor:	Prof. Dr. A.J. Cabo
Second Instructor:	Dr. J. Wong
Project Duration:	February, 2021 - July, 2023
Faculty:	Faculty of Electrical Engineering, Mathematics and Computer Science, Technical University of Delft

Contents

1 Introduction

In a collaboration with the Department of Education and Psychology at the Erasmus University of Rotterdam, research was conducted at PRIME (PRogramme of Innovation in Mathematics Education), resulting in certain hypotheses about the relationships between Perceived Monitoring and Scaffolding cues, Basic Needs (Autonomy, Relatedness and Competence) Satisfaction and Frustration, Autonomous and Controlled Motivation, and Academic Performance.
PRIME intends to redesign courses in order to improve study results, improve the connection between mathematics and engineering, and increase students' active participation and motivation [21].
Therefore it is interesting to know how the variables mentioned above are related to one another. In order to examine this, a survey was created at PRIME, consisting of a set of questionnaires concerning these variables (Academic Performance, however, was not part of the questionnaire, but was measured with actual study results).

Whilst there are several ways to examine these relationships based on the data from this survey, in this thesis we use Structural Equation Modelling (abbreviated SEM) in an attempt to answer the research questions about the formulated relationships.
Structural Equation Modelling can be described as a technique used in several sciences in order to analyse structural relationships between variables. It is particularly useful for the analysis of variables that are not directly observable, the so called latent variables. For this reason, it is especially popular in social sciences such as Psychology or Educational Sciences, as many constructs in these fields cannot be measured in a straightforward manner. Whilst social science students and graduates have some statistical background, it understandably pales in comparison to that of mathematical students and graduates. However, mathematical students are used to working with a different type of data that is usually very structured, directly observable and measurable in a straightforward manner. Therefore it is interesting to work with data from social sciences and apply mathematical techniques to this data.

Whereas some emphasis is placed on answering the research questions formulated, this is not the main focus of this thesis. The research questions can be viewed as a means to an end which is in this case examining the specifics of applying Structural Equation Modelling on any given data set.

This thesis can be viewed as complementary to or an enhancement of previous research done by two bachelor students using the same data set: 'Estimating links between latent variables using Structural Equation Modelling in R' by Plomp [20], and 'Bayesian Structural Equation Modelling explained and applied to Educational Science' by Brouwer [4].
In this thesis, the possibilities of Stuctural Equation Modelling are explored a bit more thoroughly than was done before. For example, the data set was cleaned and adjusted more thoroughly to meet the basic assumptions of Structural Equation Modelling to start with. An important part here is concerned with the way in which missing values are handled. Also, since the data is of ordinal nature, a different type of estimator is used compared to situations where continuous data is being handled. More emphasis was placed on modifying the structure of the expected relationships based on the goodness of fit results of the models created. The construct of Monitoring and Scaffolding was split into two separate constructs in order to obtain a better model, and finally, not just one, but many models were created for interpreting several sub parts of the data, resulting in several results with several interpretations that can then be compared to one another.

In this thesis, a more thorough explanation of SEM will be given in Chapter 2 where we also give an example of set of constructs, a formulated model, and a corresponding set of equations. We explain how a model can then be evaluated and in some cases adjusted to obtain a better goodness of fit. A brief explanation of the implementation in `R` is also given. In Chapter 3, the latent variables and hypotheses on the relationships between them, are formulated and a graphical presentation as well as the structural equations of the model are given. In Chapter 4 the entire questionnaires are given on Perceived Monitoring and Scaffolding cues, Basic Needs Satisfaction and Frustration, and finally Motivation. Chapter 5 focuses on meeting the assumptions of the model and cleaning the data set accordingly. A large part of this Chapter focuses on handling missing values appropriately. Chapter 6 provides more insight into the demographics of the population the survey was applied to. Chapter 7 starts with showing the initial results for the three formulated sub models and for the complete model in terms of the goodness of fit, after which adjustments are made in order to improve the models step by step. Comparisons are made between and interpretations are given of the results of the sub models and the ones of the complete model. The hypotheses are tested and in some cases, new models are formulated based on expected mediation effects between the latent variables. These mediation effects are of interest as they give more insight in the nature of the relationships. Chapter 8 provides a conclusion with respect to the main focus of this thesis, which is creating the models in the first place, but also on the content of the models (the survey itself). Also, a comparion is made to the results in previous research done by Brouwer [4] and Plomp [20]. Finally, points of discussion are given in Chapter 9.

2 Theory: Structural Equation Modelling

2.1 What is Structural Equation Modelling ?

Structural Equation Modelling (SEM) can be described as a diverse set of methods used by scientists in both experimental and observational research, consisting of computer algorithms and statistical methods for testing a network of relationships between variables [25]. Structural equation models are often used to assess so called 'latent' constructs: constructs that are not directly observable [10]. These latent variables are defined by the user in terms of the observed variables that are directly measurable, often by means of questionnaires or surveys.

SEM is widely used in social sciences, as this is a field in which variables of interest are often not directly measurable, but can be estimated by a set of questions. If we want to measure, for example, stress levels in an individual, we can ask them certain questions and combine these answers to get an idea of how much stress they are experiencing. In this example, 'Stress' is the so called latent variable, which can be measured by combining by multiple questions with respect to aspects of their life. The answers to these questions are the so called observed variables.
SEM can then be used to test the strength of the relationships between the measured variables and the latent ones.
Also, as is often the case, multiple latent variables are measured within a single questionnaire, after which SEM can also be used to assess the relationships between the latent variables.

Most of the time, the user of SEM already has an idea (based on a theoretical framework) about what items will load onto what latent variable. This is, after all, often how the questions were constructed in the first place. Therefore, the user already has an idea about what the network looks like. By applying SEM on this proposed network, the goodness-of-fit of the model can be tested, and suggestions can be made to improve the model.

2.2 Specifying the model

Basically, the model consists of sets of equations that describe the relationships between all variables. In order to apply SEM, a model must be formulated first. In general, a model consists of latent variables, observed variables, and their expected relations.

The latent variables are the variables that are not directly measurable. Stress is a good example of a latent variable. This might be measured by a set of observations, such as how often a person has frowned in the last hour, or by a set of questions, such as "How often do you have the feeling you can't cope with a situation?" These sets of observations or questions are directly measurable and are therefore called the observed variables.

Most of the time, a model consists of more than just one latent variable. Also, some latent variables can be expected to be related to one another or to influence each other. Therefore, a distinction is made between exogenous and endogenous latent variables. The exogenous latent variables are the latent variables that are only expected to be influenced by observed variables. The endogenous latent variables, on the other hand, are expected to be influenced by other (exogenous or endogenous) latent variables as well as by observed variables.

An example of a simple model is given below:

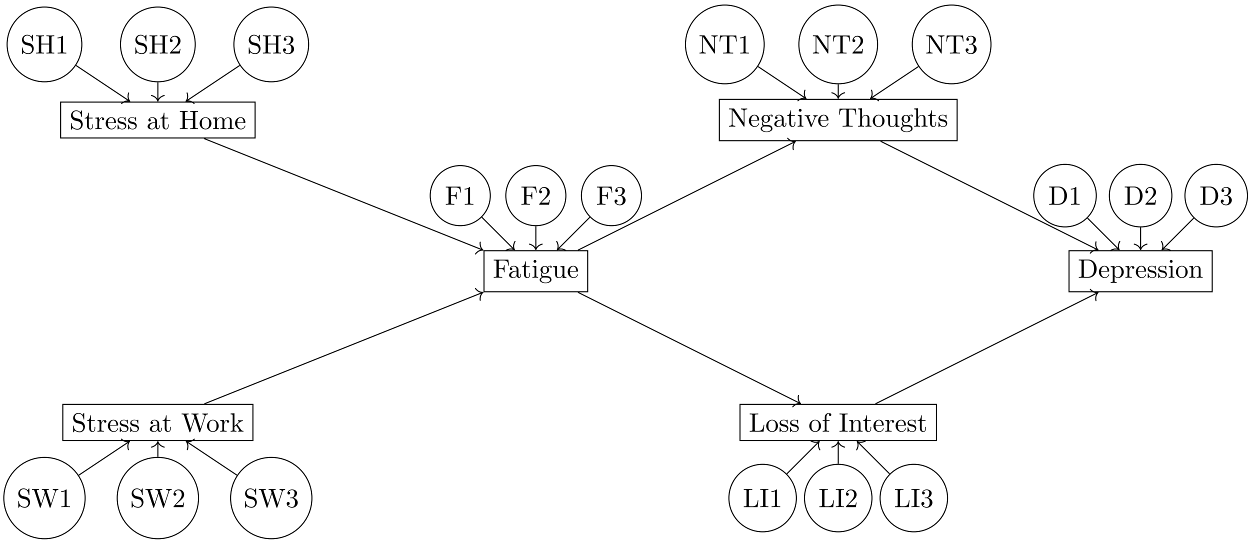

Figure 1: Example of a model

In this example, Stress at Home (SH) and Stress at Work (SW) are two exogenous latent variables, each measured by three observed variables. The other four latent variables (F=Fatigue, NT= Negative Thoughts, LI= Loss of Interest and D=Depression) are endogenous, as they are influenced by other (exogenous or endogenous) latent variables. The endogenous latent variables are in this case also measured by three observed variables each.

Note that this example is not based on any scientific research, and is simply an example made up by the author for illustration purposes.

2.2.1 The sets of equations

In SEM, two types of equations can be distinguished: the measurement equations and the structural equations.

The measurement equations relate the observed variables to the exogenous latent variables or the endogenous latent variables. Let ξ be vector of exogenous latent variables and X a vector of measurements, the observed variables that are expected to be related to the elements of ξ. We can define X in terms of ξ with the following equations:

$$X = \Delta_X \xi + \delta \tag{1}$$

where Δ_X is a matrix of coefficients explaining how each of the latent variables in ξ is described in terms of the observed variables in X. Here, δ is an error term and all δ_i are assumed to be independent of each other and identically distributed.
Now, let η be a vector of endogenous latent variables and Y a vector of measurements or observed variables that are expected to be related to the variables in η. We can define Y in terms of η with the following equations:

$$Y = \Delta_Y \eta + \varepsilon \tag{2}$$

where Δ_Y is a matrix of coefficients explaining how each of the latent variables in η is described in terms of the observed variables in Y. Here, ε is another error term, where $_i$ are assumed to be independent of each other for all i and are identically distributed.

Note that we could also choose to have $Y = X$ and let this vector consist of all measurements in the model, but this would result in larger matrices for $\Delta_X = \Delta_Y$ as they would include many zeroes for those measurements that are not related to the endogenous latent variables or the exogenous latent variables, respectively. Therefore, we choose to distinguish between X and Y.

Besides the measurement equations, the model also consists of structural equations. The structural equations are of most interest, as these are the equations that explain how the latent variables are related to each other. They are defined as follows:

$$\eta = B\eta + \Gamma\xi + \zeta \tag{3}$$

where B is a matrix explaining how the endogenous latent variables are related to the other endogenous latent variables, and Γ is a matrix explaining how the endogenous latent variables are related to the exogenous latent variables. Here, ζ is an another error term, for which ζ_i is assumed to be independent of ζ_j for all i, j and they are identically distributed.

Together, equations 1, 2 and 3 form the general model.

In order to illustrate these equations, the example from Figure 1 can be written in matrix notation as follows:

$$\begin{pmatrix} SH1 \\ SH2 \\ SH3 \\ SW1 \\ SW1 \\ SW3 \end{pmatrix} = \begin{pmatrix} \Delta_{X11} & 0 \\ \Delta_{X21} & 0 \\ \Delta_{X31} & 0 \\ 0 & \Delta_{X42} \\ 0 & \Delta_{X52} \\ 0 & \Delta_{X62} \end{pmatrix} \begin{pmatrix} SH \\ SW \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{pmatrix} \tag{4}$$

corresponds to equation 1,

$$\begin{pmatrix} F1 \\ F2 \\ F3 \\ NT1 \\ NT2 \\ NT3 \\ LI1 \\ LI2 \\ LI3 \\ D1 \\ D2 \\ D3 \end{pmatrix} = \begin{pmatrix} \Delta_{Y11} & 0 & 0 & 0 \\ \Delta_{Y21} & 0 & 0 & 0 \\ \Delta_{Y31} & 0 & 0 & 0 \\ 0 & \Delta_{Y24} & 0 & 0 \\ 0 & \Delta_{Y25} & 0 & 0 \\ 0 & \Delta_{Y26} & 0 & 0 \\ 0 & 0 & \Delta_{Y37} & 0 \\ 0 & 0 & \Delta_{Y38} & 0 \\ 0 & 0 & \Delta_{Y39} & 0 \\ 0 & 0 & 0 & \Delta_{Y4,10} \\ 0 & 0 & 0 & \Delta_{Y4,11} \\ 0 & 0 & 0 & \Delta_{Y4,12} \end{pmatrix} \begin{pmatrix} F \\ NT \\ LI \\ D \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \\ \varepsilon_{11} \\ \varepsilon_{12} \end{pmatrix} \tag{5}$$

corresponds to equation 2,
and finally

$$\begin{pmatrix} F \\ NT \\ LI \\ D \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ B_{21} & 0 & 0 & 0 \\ B_{31} & 0 & 0 & 0 \\ 0 & B_{42} & B_{43} & 0 \end{pmatrix} \begin{pmatrix} F \\ NT \\ LI \\ D \end{pmatrix} + \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} SH \\ SW \end{pmatrix} + \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \end{pmatrix} \tag{6}$$

corresponds to the most important equation, the structural equation as stated in 3.

Note that the position of the zeroes in the matrices in equation 6 depends on how the model is specified. If we add, for example, an arrow between Fatigue and Depression, B_{41} would not be 0.

From this matrix notation it becomes clear that it is actually the parameters in the matrices $\Delta_{Xij}, \Delta_{Ykl}, B_{mn}$ and Γ_{op} that we aim to find.

As mentioned before, the error terms in δ, ε, and ζ are assumed to be independent and identically distributed, and it is also assumed that $\delta_i, \varepsilon j$ and ζ_k are independent of each other for all i, j, k [27].

2.2.2 Mediation effects in SEM

Depending on the research question or hypothesis, it is sometimes important to consider mediation effects of latent variables. In words, a mediation model is used to obtain and explain the underlying mechanism of an observed relationship between a dependent and an independent variable by including a third explanatory variable: the mediator variable [16]. However, the hypothesis is not about a causal relationship between the dependent and independent variable, but it assumes that the independent variable has a cause on the mediator variable, which then results in the dependent variable. Therefore, one could say that the mediator variable is an explanation of the nature of the relationship between the dependent variable and the independent variable.

If we again look at the previous example in Figure 1 we might for example want to know whether the effects of Stress at Home (X) on Negative Thoughts (Y) are mediated (explained) by Fatigue (M). A graphical presentation of this is shown in Figure 2:

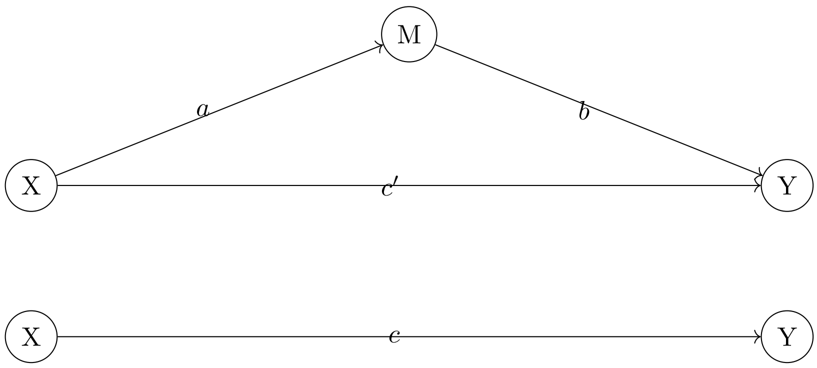

Figure 2: Graphical presentation of a general mediation model

In a mediation model of this form, the $a-$coefficient represents the effect of the independent variable X on the mediator variable M. Similarly, the $b-$coefficient represents the effect of the mediator variable M on the dependent variable Y. The difference, however, is that b was adjusted for the independent variable X. This means that given the effect a of X on M, the effect of M on Y is b.
Finally, the $c'-$coefficient represents the direct effect of X on Y, whereas the product of a and b is called the indirect effect of X on Y.

The sum of ab and c' is equal to c, which we call the total effect of X on Y:

$$c = ab + c'$$

In the example hypothesis "The effect of Stress at Home (X) on Negative Thoughts (Y) is mediated (explained) by Fatigue (M)", we want to know whether the relationship between Stress at Home and Negative Thoughts has been significantly reduced after inclusion of the mediator variable Fatigue, in which case we speak of a mediation effect.

There are three possibilities:

1. If we find that $a > 0$ and $b > 0$, but $c' = 0$, we conclude that there is a full mediation effect: X has no effect on Y when we take the mediator variable M into account.

2. We can speak of a partial mediation effect if $a > 0$, $b > 0$ and $c' > 0$, but $c' < c$. In this case, we see that we have a mediation effect $ab > 0$ but also still a direct effect $c' > 0$. The relationship between X and Y is reduced by including the mediator, but the mediator does not explain everything.

3. Finally, there is no mediation effect at all when $|c' - c| \approx 0$. (the absolute difference between c' and c is not significant) In this case, $ab \approx 0$.

Also, a distinction can be made between a positive and a negative mediation effect. If we have $c' < c$, we say there is a positive mediation effect: the mediated or indirect effect ab is positive. If a situation occurs where $c' > c$, such that we have $ab < 0$, we call this a negative mediation effect. This could occur if we find that, for example, Stress at Home attributes to someones Fatigue (there is a positive effect) but because of the Fatigue, someone has less negative thoughts (a negative effect). In this case, the direct effect could be larger than the total effect.
Note that, in the explanation above, it is assumed that c' and c are positive.

2.3 Identifying the model

The next step is to make sure that the model is identified. We say that a model is identified if there exists at least one solution for all of the model's parameters [11]. A minimum condition of identifiability is that the number of known values must equal or exceed the number of free parameters in the model . We call this the $t-$rule:

$$\frac{1}{2}p(p+1) \geq 2p + z \tag{7}$$

where p is equal to the number of measured variables and z equals the number of relations between the latent variables.
Note that the left-hand side $\frac{1}{2}p(p+1)$ equals the number of covariances (the number of elements in the covariance matrix minus the redundant elements above the diagonal (or below, as the matrix is symmetrical)). The right-hand side equals the number of free parameters or unknowns. We know from linear algebra that there can be no solution, exactly one solution, or infinitely many solutions. If the $t-$rule is not met, no solution exists for the equations. In this case we say the model is not identified; if a unique solution exists, equation 7 becomes an equality, but most often in SEM the model is over-identified (there is a strict inequality), as there are more knowns than there are free parameters. In this case there are infinitely many solutions.
It is our goal to find the 'best' solution out of these solutions by minimizing the error terms.

In our example shown in Figure 1 we have 18 measured variables and 6 relationships, and therefore the $t - rule$ holds:

$$\frac{1}{2}(18)(19) = 171 \geq 42 = 2(18) + 6$$

and we see the model is over-identified.

2.4 Estimating the model

The next step is to estimate the free parameters in the model. The most commonly used estimation method is that of Maximum Likelihood. However, in SEM, one of the assumptions of this estimator is that the data is normally distributed. (More about this and other assumptions will be discussed in Section 5.1).
We know beforehand that our data is not normally distributed, as our data is ordinal, and therefore we needed to look for other options.

Even though our data is ordinal, in SEM analysis, it is often assumed that a continuous variable underlies an ordered categorical variable [19]. This underlying continuous variable is categorized into the ordered categorical variable. Under this assumption, the measure of association that is of interest in the Modelling is the correlation between the underlying continuous variables, which is termed the polychoric correlation.
For ordered or categorical data, Diagonally Weighted Least Squares (DWLS) estimation makes use of

these polychoric correlations and has become the most popular method to obtain the parameters we are interested in using SEM[22].

Weighted Least Squares can be seen as a specialization of Generalized Least Squares, in which knowledge of the variance of observations is incorporated into the regression. The general form of a fit function can be written as

$$\hat{F} = (S - \hat{\Sigma}(\theta))^T W^{-1} (S - \hat{\Sigma}(\theta)) \tag{8}$$

Here, S and $\hat{\Sigma}(\theta)$ are correlation vectors of the data and the implied model respectively, θ is a vector of parameters, and W is a weight matrix with $\frac{1}{2}p(p+1)$ rows and columns that is specific to the estimation method [8].
The fit function characterizes the discrepancy between $\hat{\Sigma}(\theta)$ and S . The vector θ is estimated by minimizing F.

If we take $W = n \cdot \mathrm{Diag}(\Gamma)$, where n is the sample size, Γ is the asymptotic variance and covariance matrix of polychoric correlations, and $\mathrm{diag}(\Gamma)$ is a diagonal matrix with all the diagonal elements the same as the diagonal elements in Γ, F becomes the Diagonally Weighted Least Squares fit function. The detailed calculation of $n \cdot \mathrm{Diag}(\Gamma)$, is described in [19].

2.5 Evaluating the model

In applications of structural equation Modelling, a critical step is to evaluate the goodness-of-fit of the proposed model with the data. In order to test the goodness-of-fit of the model, several mathematical measures are used.

The chi-square statistic is used to perform a test of perfect model fit. It is essentially a measure of deviance between the model-implied covariance matrix, and the observed covariance matrix [6].

Definition 2.5.1. *The Chi-Square (χ^2) statistic is defined as*

$$\chi^2_{df} = \sum_i \frac{(O_i - E_i)^2}{E_i} \tag{9}$$

where O_i are the observed values and E_i are the expected values.

The chi-square test is a test of statistical significance where the chi-square value and the degrees of freedom are used to calculate a $p-$value. The null hypothesis that the predicted model and observed data are equal is tested. Because you want your predictions to match the actual data as closely as possible, you do not want to reject this null hypothesis, meaning that a good model fit is indicated by a non-significant result for this test.

In our case, the χ^2 statistic is not very reliable, however.
This mostly has to do with the fact that the chi-square test is very sensitive to sample size [2],[9]. The larger the sample size, the greater the chances of obtaining a statistically significant chi-square. And given that it is agreed upon that SEM should only be conducted with large sample sizes, the chi-square test is almost guaranteed to be significant, even at higher significance cutoffs (e.g. 0.001). This turns out to be the case even more when ordinal data is used instead of continuous normally distributed data [11]. For these reasons this statistic does not provide any useful information, and other measures of fit need to be considered.

Alternative goodness-of-fit measures have been developed in an attempt to provide additional information about the usefulness of the hypothesized model [24]. Many of these fit indices are developed based on the chi-square test, and are based on a comparison with a so called 'null model' that assumes there are no correlations between the variables at all.
One needs to keep in mind, however, that the issues with the Chi-square test are not entirely resolved by

looking at these alternative measures, as they still involve the Chi-square statistic.

Definition 2.5.2. *The Bentler-Bonett Index [3] or Normed Fit Index (NFI) is an incremental measure of fit defined as*

$$NFI = \frac{\chi^2_N - \chi^2_M}{\chi^2_N}$$

where N is the null model, and M is the model as proposed by the user.

The null model can be interpreted as the 'independence model', where the covariances are assumed to be zero in the model.
The Bentler-Bonett Index looks at how much the χ^2-value has improved with respect to the null model. However, there is no correction for the number of parameters. Therefore, an improved version of the Bentler-Bonett Index is the Tucker-Lewis Index (TLI)[3], [26], measuring a relative reduction in misfit compared to the Null model, taking into account the degrees of freedom:

Definition 2.5.3. *Let F_M be the the minimized fit function of the proposed model, and F_N the minimized fit function of the Null model, at the population level. The unscaled Tucker Lewis Index is defined as*

$$\begin{aligned} TLI_U &= 1 - \frac{\frac{F_M}{df_M}}{\frac{F_N}{df_N}} && (10) \\ &= 1 - \frac{F_M \cdot df_N}{F_N \cdot df_M} && (11) \end{aligned}$$

Here, the fit function is as described in equation 8.

In modern software like `R`, however, the scaled version of the TLI is used, estimating parameters α, a scaling parameter, and β, a shifting parameter, for an estimation of the sample TLI [30]:

$$\text{TLI}_S = 1 - \frac{\hat{\alpha}_M (n-1)\hat{F}_M + \hat{\beta}_M - df_M}{\hat{\alpha}_N (n-1)\hat{F}_N + \hat{\beta}_N - df_N} \cdot \frac{df_N}{df_M} \tag{12}$$

Here, n refers to the sample size.

In general, TLI ≥ 0.95 is a commonly used cutoff criterion for the goodness-of-fit [9],[28].

Since both the NFI and the TLI are comparisons with the Null model, we expect high values of NFI and TLI in case we have high correlations in our proposed model, since this increases the χ^2-value and therefore also the $\frac{\chi^2}{df}$-value of the proposed model.

Very similar to the TLI, we also have the CFI:

Definition 2.5.4. *Let F_M be the the minimized fit function of the proposed model, and F_N the minimized fit function of the Null model, at the population level. The unscaled Comparative Fit Index (CFI) is defined as*

$$CFI_U = 1 - \frac{F_M}{F_N} \tag{13}$$

Again, the fit function is as described in equation 8.

Note that the CFI is very similar to the TLI, with the only difference that there is no correction for the degrees of freedom. As with the TLI, the scaled version of the CFI is used in `R`, estimating parameters α, a scaling parameter, and β, a shifting parameter, for an estimation of the sample CFI [30]:

$$\text{CFI}_S = 1 - \frac{\hat{\alpha}_M (n-1)\hat{F}_M + \hat{\beta}_M - df_M}{\hat{\alpha}_N (n-1)\hat{F}_N + \hat{\beta}_N - df_N} \tag{14}$$

Again, n refers to the sample size.

Since we have that $df_N < df_M$ (correlations are fixed to 0 in the Null model), and therefore $\frac{df_N}{df_M} < 1$, we always have CFI > TLI.
As the CFI ranges between 0 and 1, it is considered a normed fit index, with higher values indicating a better fit. The most commonly used criterion for a good fit is CFI ≥ 0.95 [28].
The way $\hat{\alpha}$ and $\hat{\beta}$ are estimated is described in detail in [1] and is beyond the scope of this thesis.

Whereas the TLI and the CFI both compare the proposed model to the Null model, there are also fit measures that do not make such a comparison. An commonly used measure of fit is the RMSEA [5].

Definition 2.5.5. *The unscaled Root Mean Square Error of Approximation (RMSEA) can be defined as*

$$RMSEA_U = \sqrt{\frac{F_M}{df_M}} \tag{15}$$

at the population level.

Here, we have that F_M is the minimized fit function as described in equation 8. As the fit function was minimized, this indicates that a lower value of the RMSEA indicates a better fit.

As is the case with the TLI and CFI, the RMSEA is also usually scaled in modern software like `R` [30]. Therefore, the scaled function is normally used:

$$\text{RMSEA}_S = \sqrt{\frac{\hat{\alpha}_M(n-1)\hat{F}_M + \hat{\beta}_M}{(n-1)df_M} - \frac{1}{n-1}} \tag{16}$$

The RMSEA is usually reported with a confidence interval. An RMSEA ≤ 0.06 is considered to indicate a good fit [9]. On the other hand, an RMSEA of ≥ 0.10 is indicative of a very bad model [5].

Another absolute measure of fit, this time without a penalty for the number of parameters, is the SRMR:

Definition 2.5.6. *The (Standardized) Root Mean Square Residual (SRMR) is an absolute measure of misfit and is defined as the standardized difference between the residuals of the sample covariance matrix and the hypothesized model, and is quantified by*

$$SRMR = \sqrt{\frac{2\Sigma_{i,j}\left(\frac{\hat{\Sigma}(\theta)_{i,j} - S_{i,j}}{S_{i,i}S_{j,j}}\right)^2}{p(p+1)}} \tag{17}$$

where we sum over all rows $i \in (1, ..., k)$ and columns $j \in (1, ..., l)$ of the model-implied covariance matrix $\hat{\Sigma}(\theta)_{i,j}$ and the sample covariance matrix $S_{i,j}$.

(as with the RMSEA, lower values indicate a better fit) The RMR and SRMR provide estimates of the average misfit for each estimated versus observed variance/covariance parameter. It quantifies the square root of the average of the squared residuals comparing the model-implied covariance matrix and observed covariance matrix.
The SRMR ranges between 0 and 1, thus it is standardized, where a value 1 indicates that the residuals of the elements of the model-induced covariance matrix are on average as large as the elements of the sample covariance matrix. Hence, a value close to 0 is desirable.
A good fit is obtained when the SRMR ≤ 0.05, but a fair fit is obtained when the SRMR ≤ 0.08.[9] [18].
A value ≥ 0.1 is considered poor.

In this research, all models will be tested for goodness-of-fit. The following table provides an overview of how each of these values can be interpreted, based on a combination of research mentioned above.
Note that the cut-off values in table 1 are commonly used in Structural Equation Modelling, but they are largely based on intuition and experience rather than on statistical justification [17]. However, this table provides an anchor we can use to interpret our results.

Goodness-of-fit measure	perfect	good	acceptable	bad
CFI		≥ 0.95	≥ 0.9	
TLI		≥ 0.95	≥ 0.9	
RMSEA	0	≤ 0.05	≤ 0.06	≥ 0.1
SRMR	0	≤ 0.05	≤ 0.08	≥ 0.1

Table 1: Interpretation of goodness-of-fit measures

2.6 Modifying and improving the model

If the values discussed in the previous Chapter are not what is considered good enough, a modification to the model can be made in order to obtain a better fit [15]. Modification Indices are a way of improving the model by identifying parameters which, if included, would improve model fit. However, it is important that the use of Modification Indices should be informed by theory, as Modification Indices may suggest paths that don't make theoretical sense. If paths are added to the model based only on the data, overfitting is likely to occur. Therefore, we need to consider carefully which paths to add to the model and which paths we should leave out.

To discover what parameters are most useful additions to our model, we consult `R`. This will be explained more thoroughy in the next section.
Another way of improving the model might be to remove some parameters. This can be done by manually using the SRMR from equation 17 to discover in which cases the misfit is reduced. If it turns out that the SRMR is much lower when a certain item is removed, it might be a good idea to delete this item from the survey entirely. Again, it is important that we always consider the content of the item before we remove it, as it has to make sense theoretically as well.

2.7 Structural Equation Modelling in `R`

In general, we transform the measurement and structural equations into a model. Recall equations 1, 2 and 3:

$$\begin{aligned} X &= \Delta_X \xi + \delta \\ Y &= \Delta_Y \eta + \varepsilon \\ \eta &= B\eta + \Gamma\xi + \zeta \end{aligned}$$

Based on our expectations of what these equations should look like, we tell `R` what measured variables are related to what latent variables, and what latent variables are related to each other. We do not need to define the exact relationships (i.e. the expected coefficients of the matrices). These are calculated by `R`.
We use `R` to test our proposed model, and give us the coefficients for the matrices Δ_X, Δ_Y, B and Γ. When the model is tested, an output is generated, producing among others the goodness-of-fit measures described in Section 2.5 and the coefficients for the matrices. Based on this output, we can redefine our model if the goodness-of-fit measures turn out to be very bad, or we can adjust the model slightly, either by adding regressions between observed variables, or by removing some observed variables.

In `R`, a specific package for Latent Variable Analysis (LaVaAn, in `R: lavaan`) is used.
Special attention needs to be given to some matters concerning our specific topic:

1. First, we are dealing with ordinal data. The specific nature of our data is discussed in Chapter 4. The fact that our data is ordinal, needs to be specified in `lavaan`, as one of the assumptions of the standard method is that the data is continuous. Without specifying that the data is ordinal, the standard Maximum Likelihood estimator is used to compute the coefficients of the matrices. In our case, this is not the right estimator: we need to use the so called Diagonally Weighted Least Squares (DWLS) estimator. The DWLS estimator yields more accurate factor loading estimates than the maximum likelihood estimator when data is categorical or ordinal, as it incorporates extra

(non-negative) weights associated with each data point, into the fitting criterion. [18].
The DWLS estimator is used automatically when we include the simple command "`ordered=TRUE`" to the model.

2. Also, we need to specify how missing values are handled. In our data set, we are dealing with a lot of missing values. Most of them could be dealt with by inspecting the data manually, as will be explained in Section 5.3. The much smaller number of missing values that remain, can be dealt with by `R`, and we let it handle them in a pairwise manner by specifying the command "`missing=''pairwise''`." Handling missing values by pairwise deletion allows us to use more data as opposed to when listwise deletion is used. With listwise deletion, whenever `R` detects there is a missing value, the entire subject (student) is removed from the analysis. With pairwise deletion, however, only the variable with missing values is removed, but all other variables are still used for this subject.

3. Sometimes `R` has trouble computing standard errors, even though it is not always clear why. As this was the case in some of our models as well, bootstrapping was used by applying the command "`se=''boot`", in order to compute standard errors. The bootstrap method was applied only to those models where a problem occurred with computing standard errors, not to all models.
When bootstrapping was used, 1000 samples were taken from the data in order to produce standard errors for the estimates. This had to be done, because without standard errors it is impossible to know whether the regressions obtained are statistically significant, in which case no meaningful interpretations can be given to the regressions obtained. Some attention will be given to this shortcoming in the Discussion in Chapter 9.

4. Often, the initially proposed model does not meet the goodness-of-fit requirements right away, which also turned out to be the case in most of our models. Therefore, modifications to the models needed to be made in order to improve these measures. In order to know what modifications could be useful, Modification Indices were used. These are obtained in `R` by using the function "`modindices(fit, sort=TRUE)`." The output of this function is then a sorted list of Modification Indices per parameter. A higher Modification Index suggests a better improvement to the model if this specific parameter is added to the model. The parameter on top of the list is, since we sorted the list, the most useful parameter to add to the model. However, the addition of the parameter needs to make sense theoretically and should therefore be manually checked. By adding the suggested parameters to the model (which are all regressions in our case), we allow the regressions between survey items not to be fixed to 0.
There is, however, a risk of overfitting. After all, we could choose to add all possible regressions to our model in which case we obtain the best goodness-of-fit measures possible. The Modification Index is the χ_1^2 value, by which the model fit would improve if a particular path was added. Values bigger than 3.84 indicate that the model would be 'improved', and the $p-$value for the added parameter would be < 0.05, and values larger than 10.83 indicate the parameter would have a $p-$value < 0.001 [29]. In this research, we therefore decided to add only those regressions with a MI larger than 10.83, with the extra condition that they make sense theoretically. The process of adding parameters to our models is described in detail in Section 7.2.

2.8 Interpreting the results

Results will be obtained, mainly, in the form of the goodness-of-fit measures we discussed earlier: the CFI, TLI, RMSEA and SRMR. These measures, as explained before, indicate whether the model is a good fit for the data.
Once we have these results, it is also interesting to see how each of the variables contribute to the model. For this, we look at the parameters, or coefficients, that describe the relationships between the latent variables. Finally, we aim to test a set of hypotheses and consider the conclusion with respect to these hypotheses a result as well.

3 The Model

3.1 Latent variables used in the model

This research revolves around the relationship between 11 latent variables:

1. Monitoring (Mon), expected to be measured by part of the survey as described in Chapter 4.1
2. Scaffolding (Scaf), expected to be measured by part of the survey as described in Chapter 4.1
3. Autonomy Satisfaction (AS), expected to be measured by part of the survey as described in Chapter 4.2
4. Autonomy Frustration (AF), expected to be measured by part of the survey as described in Chapter 4.2
5. Relatedness Satisfaction (RS), expected to be measured by part of the survey as described in Chapter 4.2
6. Relatedness Frustration (RF), expected to be measured by part of the survey as described in Chapter 4.2
7. Competence Satisfaction (CS), expected to be measured by part of the survey as described in Chapter 4.2
8. Competence Frustration (CF), expected to be measured by part of the survey as described in Chapter 4.2
9. Autonomous Motivation (AM), expected to be measured by part of the survey as described in Chapter 4.3
10. Controlled Motivation (CM), expected to be measured by part of the survey as described in Chapter 4.3
11. Academic Performance (AP), measured by the final grade for the course the survey was about: Calculus (CSE1200) at the Bachelor program 'Computer Science and Engineering' at TU Delft.

Note that the variables Monitoring and Scaffolding are measured by one part of the survey and together are called 'Monitoring and Scaffolding Cues' or 'MSC.' Similarly, Autonomy Satisfaction, Autonomy Frustration, Relatedness Satisfaction, Relatedness Frustration, Competence Satisfaction and Competence Frustration are referred to as the 'Needs' and are also part of the same questionnaire. Finally, the Motivation part of the survey measures the two types of Motivation: Autonomous and Controlled Motivation.

3.2 Hypotheses

In order to answer the research question

"**What are the relationships between student perceptions of teacher cues in Monitoring and Scaffolding, Satisfaction and Frustration of Needs for Relatedness and Competence, Motivation, drop-outs, and Academic performance?**"

Among others, the following hypotheses were formed:

1. Perceived teacher Scaffolding and Monitoring cues are positively related to Competence and Relatedness Satisfaction and negatively related to Competence and Relatedness Frustration
2. Autonomy Satisfaction is positively related to Autonomous Motivation and negatively to Controlled Motivation

3. Autonomy Frustration is positively related to Controlled Motivation and negatively to Autonomous Motivation
4. Competence and Relatedness Satisfaction is negatively related to Controlled Motivation
5. Competence and Relatedness Frustration are negatively related to Autonomous Motivation, but less strongly than the relationships between Competence and Relatedness Satisfaction and Autonomous Motivation, and Competence and Relatedness Frustration and Controlled Motivation
6. Competence and Relatedness Satisfaction positively mediate the relationship of perceived teacher cues and Autonomous Motivation
7. Competence and Relatedness Frustration negatively mediate the relationship of perceived teacher cues and Controlled Motivation
8. there is a positive relationship between Autonomous Motivation and performance and a negative, but less strong, relationship between Controlled Motivation and performance

3.3 Graphical presentation of the model

Based on some hypotheses formulated in Chapter 3.2, the following graph can be drawn to present our expectation of what the model should look like:

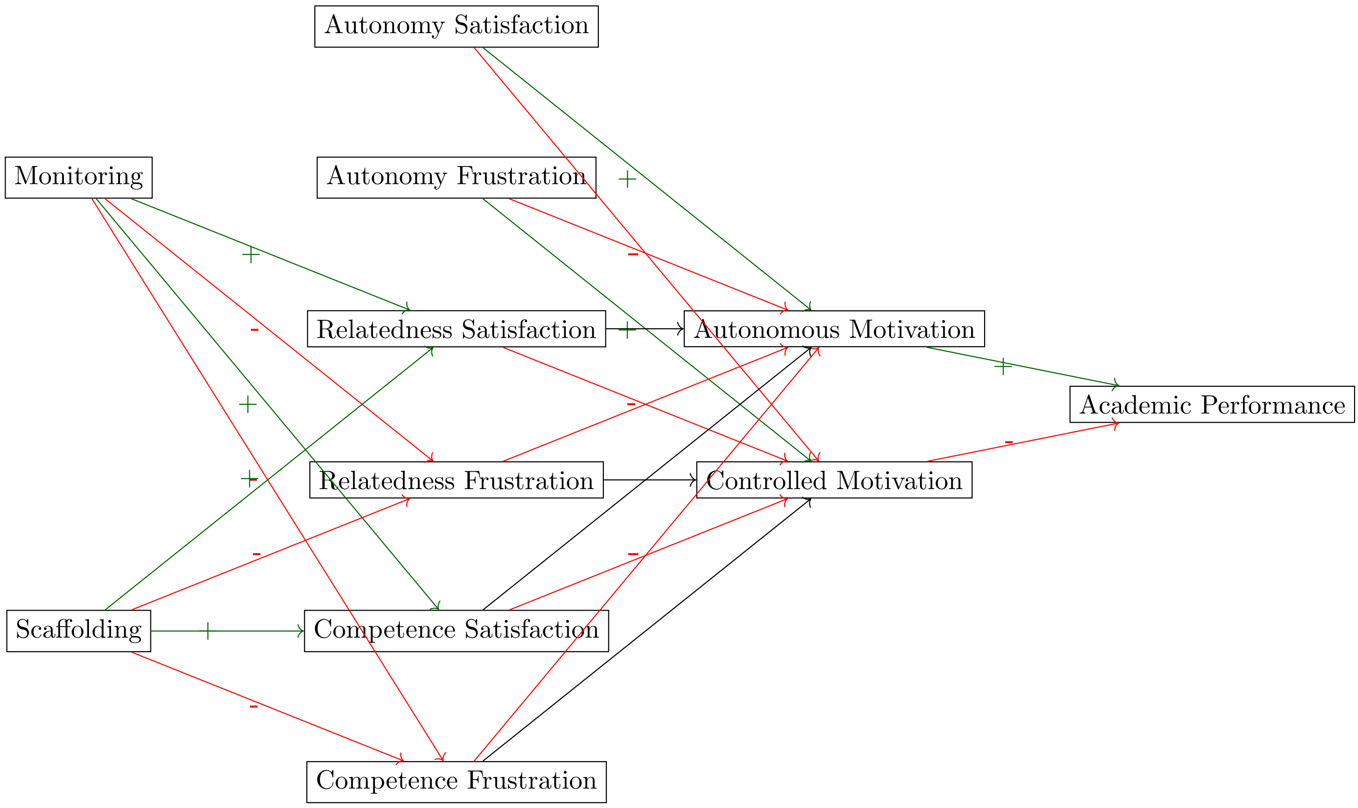

Figure 3: Graphical Presentation of the hypothesized model

Positive relationships are colored green, whereas negative relationships are colored red. With respect to the fifth hypothesis, it is unclear whether some of the relationships are expected to be positive or negative, so these arrows are colored black.
Note that some of the hypotheses involve the strengths of the relationships, mediation effects or moderation effects. These are not represented in this graph.
Also, note that this Figure represents the total model. In the next Chapter, we will see that it can be convenient to split up the model into three parts.

3.4 Structural equations for the model

The model presented in Figure 3 can be presented in matrix notation as follows:

$$\begin{pmatrix} RS \\ RF \\ CS \\ CF \\ AM \\ CM \\ AP \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{5,1} & \beta_{5,2} & \beta_{5,3} & \beta_{5,4} & 0 & 0 & 0 \\ \beta_{6,1} & \beta_{6,2} & \beta_{6,3} & \beta_{6,4} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta_{7,5} & \beta_{7,6} & 0 \end{pmatrix} \begin{pmatrix} RS \\ RF \\ CS \\ CF \\ AM \\ CM \\ AP \end{pmatrix}$$
$$+ \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & 0 & 0 \\ \gamma_{2,1} & \gamma_{2,2} & 0 & 0 \\ \gamma_{3,1} & \gamma_{3,2} & 0 & 0 \\ \gamma_{4,1} & \gamma_{4,2} & 0 & 0 \\ 0 & 0 & \gamma_{5,3} & \gamma_{5,4} \\ 0 & 0 & \gamma_{6,3} & \gamma_{6,4} \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \text{Mon} \\ \text{Scaf} \\ AS \\ AF \end{pmatrix} + \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \\ \zeta_5 \\ \zeta_6 \\ \zeta_7 \end{pmatrix}$$

Again, this set of equations represents the total model. For the three separate models discussed in the next chapter, similar, somewhat more simplistic sets of equations can be set up.

3.5 Programming the model

A complete overview of how the model was programmed in `R` is given in the appendix in section 9.

4 The Survey

The analyses in this thesis are based on the following survey, consisting of three parts:

4.1 Perceived Monitoring and Scaffolding cues

The first part of the survey consists of 30 questions regarding perceived Monitoring and Scaffolding cues. Perceived Monitoring and Scaffolding cues were measured with an adapted version of the Student Assessment for Learning Questionnaire (SAFL-Q; Pat-EL, 2013) which has been used by Cabo and Klaassen (2019). The questionnaire comprises sixteen Monitoring items and fourteen Scaffolding items which are presented as statements. Participants were asked to indicate on a Likert scale whether they agreed (1 = strongly disagree, 5 = strongly agree).

1. The lecturer encourages me to reflect on how I can improve on my assignments.
2. After examining the test results, the lecturer discusses the answers given to the test in class.
3. Whilst working on my assignments, the lecturer asks how I think I am doing.
4. The lecturer stimulates us to think about what we want to learn in university.
5. The lecturer gives me the opportunity to decide on my own learning strategies.
6. The lecturer inquires about what went well and what went badly in my work.
7. The lecturer encourages me to reflect on my learning process.
8. The lecturer stimulates me to think about how to improve next time.
9. The lecturer shows how to find my strengths concerning my study skills.
10. The lecturer shows how to identify my weaknesses concerning my study skills .
11. I am encouraged by the lecturer to improve my learning process.
12. The lecturer gives me guidance to assist my learning.
13. The lecturer discusses assignments to help me understand the subject matter better.
14. The lecturer discusses with me the progress I make.
15. After each assessment the lecturer informs us on how to improve the next time.
16. The lecturer discusses how to make full use of my study skills to improve my assignment.
17. I am aware of my weak points in the application of study skills.
18. The lecturer offers strategies to improve my study skills.
19. When I do not understand a topic, the lecturer tries to explain it in a different way.
20. The lecturer provides me with hints to help understand the subject matter.
21. During class I have an opportunity to show or share what I have learned.
22. The lecturer asks questions in a way I understand.
23. The lecturer asks questions that help me gain understanding of the subject matter.
24. The lecturer allows for my contribution during the lesson.
25. I have the opportunity to ask my classmates questions during class.
26. The lecturer makes me aware of the areas I need to work on to improve my results.

27. There is an opportunity to ask questions during class.
28. I am aware of the criteria by which my assignment will be evaluated.
29. When I receive an assignment it is clear to me what I can learn from it.
30. The assignments allow me to show what I am capable of.

Items 1-16 are expected to measure the latent variable 'Monitoring,', whereas items 17-30 are expected to measure the latent variable 'Scaffolding.'
In later chapters, the items are referred to as MSC1, MSC2, ..., MSC30.

4.2 Basic Needs: Satisfaction and Frustration

The basic Needs were measured with an adjusted version of the Basic Psychological Need Satisfaction and Frustration Scale (BPNSNF; Chen et al., 2015). This questionnaire consists of 24 statements for which participants can indicate the degree to which the statements are true for them at this point in their life on a Likert scale of 1 (not true at all) to 5 (completely true). The 24 statements are divided into 6 subscales which contain 4 items each: Relatedness Satisfaction (RS) and Frustration (RF), Competence Satisfaction (CS) and Frustration (CF), and Autonomy Satisfaction (AS) and Frustration (AF):

In the context of this course..

1. I feel a sense of choice and freedom in the things I undertake.
2. Most of the things I do feel like “I have to”.
3. I feel that the fellow students and lecturers I care about also care about me.
4. I feel excluded from the group I want to belong to.
5. I feel confident that I can do things well.
6. I have serious doubts about whether I can do things well.
7. I feel that my decisions about course activities reflect what I really want.
8. I feel forced to do many things I wouldn't choose to do.
9. I feel connected with fellow students and lecturers who care for me, and for whom I care.
10. I feel that fellow students and lecturers who are important to me are cold and distant towards me.
11. I feel capable at what I do.
12. I feel disappointed with many of my performance.
13. I feel my choices about course activities express who I really am.
14. I feel pressured to do too many things
15. I feel close and connected with other fellow students and lecturers who are important to me.
16. I have the impression that fellow students and lecturers I spend time with dislike me.
17. I feel competent to achieve my goals.
18. I feel insecure about my abilities.
19. I feel I have been doing what really interests me.
20. My daily activities feel like a chain of obligations.

21. I experience a warm feeling with fellow students and lecturers I spend time with.
22. I feel the relationships I have with my fellow students and lecturers are just superficial.
23. I feel I can successfully complete difficult tasks.
24. I feel like a failure because of the mistakes I make.

Autonomy Satisfaction is expected to be measured by items 1, 7, 13 and 19.
Autonomy Frustration is expected to be measured by items 2, 8, 14 and 20.
Relatedness Satisfaction is expected to be measured by items 3, 9, 15 and 21.
Relatedness Frustration is expected to be measured by items 4, 10, 16 and 22.
Competence Satisfaction is expected to be measured by items 5, 11, 17 and 23.
Competence Frustration is expected to be measured by items 6, 12, 18 and 24.

In later chapters, the items are referred to as Needs1, Needs2, ..., Needs24.

4.3 Motivation

This part of the questionnaire can be divided into a part concerning so called 'Autonomous Motivation,' and a part about so called 'Controlled Motivation.' Motivation was measured with an adapted version of the academic self-regulation scale (Ryan & Conell, 1989).
The items were presented as statements for which students could indicate whether they were true for them or not on a Likert scale of 1 (Completely not important) to 5 (Very important). Initially, the survey intends to measure four types of Motivation: intrinsic Motivation, introjected regulation, identified regulation and external regulation. All four were all assessed with 4 items each, leading to 16 items in total. The intrinsic Motivation and identified regulation subscales were combined to form an Autonomous Motivation part and the introjected regulation and external regulation subscales were combined to form a Controlled Motivation composite (Sheldon, Ryan, Deci, Kasser, 2004):

Why are you studying in general? I'm studying ...

1. Because I want others to think I'm a good student.
2. Because I enjoy doing it.
3. Because I'm supposed to do so.
4. Because I want to learn new things.
5. Because I would feel ashamed if I didn't study.
6. Because others (parents, friends, etc.) oblige me to do so.
7. Because it's an exciting thing to do.
8. Because it's a meaningful choice to me.
9. Because that's what others (e.g., parents, friends) force me to do.
10. Because that's what others (parents, friends, etc.) expect me to do.
11. Because I'm highly interested in doing this.
12. Because I would feel guilty if I didn't study.
13. Because it is personally important to me.
14. Because I want others to think I'm smart.

15. Because it's fun.

16. Because this is an important life goal to me.

The latent variable 'Autonomous Motivation' is expected to be measured by items 2, 4, 7, 8, 11, 13, 15 and 16, whereas the latent variable 'Controlled Motivation' is expected to be measured by items 1, 3, 5, 6, 9, 10, 12 and 14.

In later chapters, the items are referred to as M1, M2, ..., M16.

Note that all items were answered in the context of one specific course: 'Calculus (CSE1200)' for the Bachelor "Computer Science and Engineering."

5 Cleaning and Preparing the Data

Before we start with applying SEM on the data, we need to examine the data and clean/prepare it. This needs to be done because we first of all have to meet the basic assumptions of SEM that are given below:

5.1 Basic Assumptions of the Model

1. **The data is multivariate normally distributed**
   Normality of observations is the first assumption before building the model and checking its fit indices. The observations must draw from a continuous and multivariate normal population. However, in reality, normality of data is a condition that happens rarely [13].
   Whereas it is not impossible to fit a SEM to non-normal data, it can result in inflated model test statistics (for example a lower CFI or a higher RMSEA) and under-estimated standard errors when the standard Maximum Likelihood Estimator is used [7]. We are dealing with ordinal data, as subjects answered questions on a scale (as is clear from the Survey described in Chapter 4). As we discussed thoroughly in section 2.4, the Diagonally Weighted Least Squares Estimator has become the most popular estimation technique to obtain parameters in case the data is ordinal, and therefore it is still possible to use SEM on our data.

2. **There is no missing data.**
   Another assumption is that there is no missing data. However, this is a rather unrealistic assumption when human beings are used to obtain data and therefore there is almost always at least some missing data. A way to deal with this, is to test that if there is some missing data, it is missing at random (MAR) [13]. It is however problematic if the number of missing data is large.
   This assumption appears to also be violated in the original data set. By inspection, there is a large number of missing values, and also, they do not appear to be at random. We will look at this more closely in subsection 5.3.

3. **The data contains no outliers.**
   We do meet the assumption that there are no outliers. The constructs of Monitoring and Scaffolding, Basic Needs, and Autonomous/Controlled Motivation are all measured on a scale from (1) to (5), which leaves no room for outliers. Also, Academic Performance is measured by grades on a scale from (1) to (10), so the same applies here.

4. **The relationships between all constructs are linear.**
   The relationships between the observed variables and their constructs and between one construct and another is expected to be linear. This is an assumption that is not checked, but simply assumed by applying SEM. The linearity of the model becomes especially clear from the matrix notation as is shown in section 2.2.1.

5.2 The data set

In the original data set, 222 students participated in the survey. These students were selected from a Calculus course for the Bachelor "Computer Science and Engineering" at the Technical University in Delft, and were asked to fill in this survey manually by one of the lecturers. Afterwards, all data was read and collected in one large `.csv`-file.
However, not all students gave their consent to be part of the data analysis (10 students did not give consent), which left in total 212 students to work with.

In total, all students received the survey items as described in Chapter 4. Other than the questions on Monitoring and Scaffolding, Basic Needs and Motivation, however, also some demographic questions were asked. Students were asked to give their age, gender, highest level of education, country of highest education, past performance in mathematics and grade-goal for the final exam.
These are analyzed in Chapter 6.

Information about which of the lecturers the students had lessons from was also collected, as there

were eight lecturers in total and the Monitoring and Scaffolding questions specifically revolve around how the students perceive the behaviour of the lecturer.
In order to measure Academic Performance, the grades that the students achieved afterwards, were added to the data set as well. From the perspective of PRIME, it is interesting to know how all variables in the survey are related one another but especially to Academic Performance, as, based on the outcome, it may be possible to make changes in the way education is given, in order to improve Academic Performance of students.

5.3 Handling Missing Values and inconsistent data

Before we give an overview of the demographics, it is important to clean the data set. This was first of all done by inspection. It appeared to be the case that the number of missing values was very high for one of the lecturers compared to the others. Almost all of this lecturer's students had a large number of missing data:

Lecturer	A	B	C	D	E	F	G	H
# Students with consent	16	14	21	29	22	36	30	44
# Missing	1	2	220	20	51	3	0	2

Table 2: Number of Missing Values per Lecturer for 30 items on Monitoring and Scaffolding

Based on table 2 we decided to remove all students from lecturer C from the data set as the number of missing values is disproportionately large.

We then continued to look at the number of missing values per item. We decided that for Monitoring and Scaffolding, item 2 ("After examining the test results, the lecturer discusses the answers given to the test in class") resulted in a lot of missing values as well. Some students placed comments with this question, stating there had not been a test yet. Therefore, we decided to remove question 2 from the Monitoring and Scaffolding construct.
For the other items, no significantly large numbers of missing values were detected.

After examining the number of missing values per item, we also looked at the students individually. Based on this, we decided to remove five students who had between 15 to 24 missing values on the construct for Basic Needs, which consists in total of 24 items. These students were removed from the entire data set, not just from the Basic Needs construct.
For the Motivation part of the questionnaire, which consists of 16 items, there were three students with all missing values. Therefore, these students were also removed from the entire data set.
Finally, one person did not seem to have answered the questions seriously. He did not have a valid studentnumber, said he was age 65, and had repeating patterns in his answers, such as (1,2,3,4,5,4,3,2,1). This person was also removed from the data set.
When it comes to the inconsistency in the data, some decisions were made as well:

- Comments placed behind a clear response were removed to make the data set more readable. Examples of these comments are "not sure" or "between 2 and 3".
- Some students answered (2.5), (3.5) or (4.5) instead of an integer from (1) to (5). Since the answer (3) denotes a neutral answer, and there is a tendency to remain neutral, we decided to round down all answers below 3, and round up all answers above 3, to decrease the neutrality of the answers a little bit. This is a choice that was made that needs to be considered when analyzing the results.
- When answers to the demographic part of the questionnaire were unclear, they were removed and thus set to missing. Spelling errors were corrected.
- For Gender, all answers there were not a clear 'Male' or 'Female' were set to 'Other.'

The next step is to test whether the missing values that still remain, are so called 'Missing At Random (MAR).'
This was tested in several ways:

- First, a Fisher's Exact Test was performed to see whether there was a gender difference in the frequencies of the number of students with certain numbers of missing values. The result was that there is no significant difference, with $p \approx 0.35$, for the following frequencies:

# Missing Values	0	1	2	3	4	5	6	7	8	9	11	12	13	20
# Males	6	35	64	13	2	2	0	12	7	6	2	1	1	1
# Females	1	3	8	3	0	0	1	3	3	2	0	0	0	0
# Other	0	3	4	0	1	0	0	1	0	1	0	1	0	0

Table 3: Frequencies of numbers of total missing values for each gender

and $p \approx 0.41$ for the frequencies of missing values for the survey items only:

# Missing Values	0	1	2	3	4	5	10	11	20
# Males	124	17	3	4	1	0	1	1	1
# Females	18	4	1	0	0	1	0	0	0
# Other	9	1	0	0	1	0	0	0	0

Table 4: Frequencies of numbers of missing values for survey items only for each gender

Note that, even though the frequencies may seem to differ much between males and females, we need to take into account the fact that we have many more males in our student population than females. More about this will be presented in Chapter 6.

- A correlation between Age and the number of missing values was examined, too. We find Pearson's $r \approx 0.18$ for Total Missing ($p \approx 0.013$), and $r \approx -0.04$ for Survey Missing, with $p \approx 0.51$. The correlation plots (Jittered) can be seen in Figure 4:

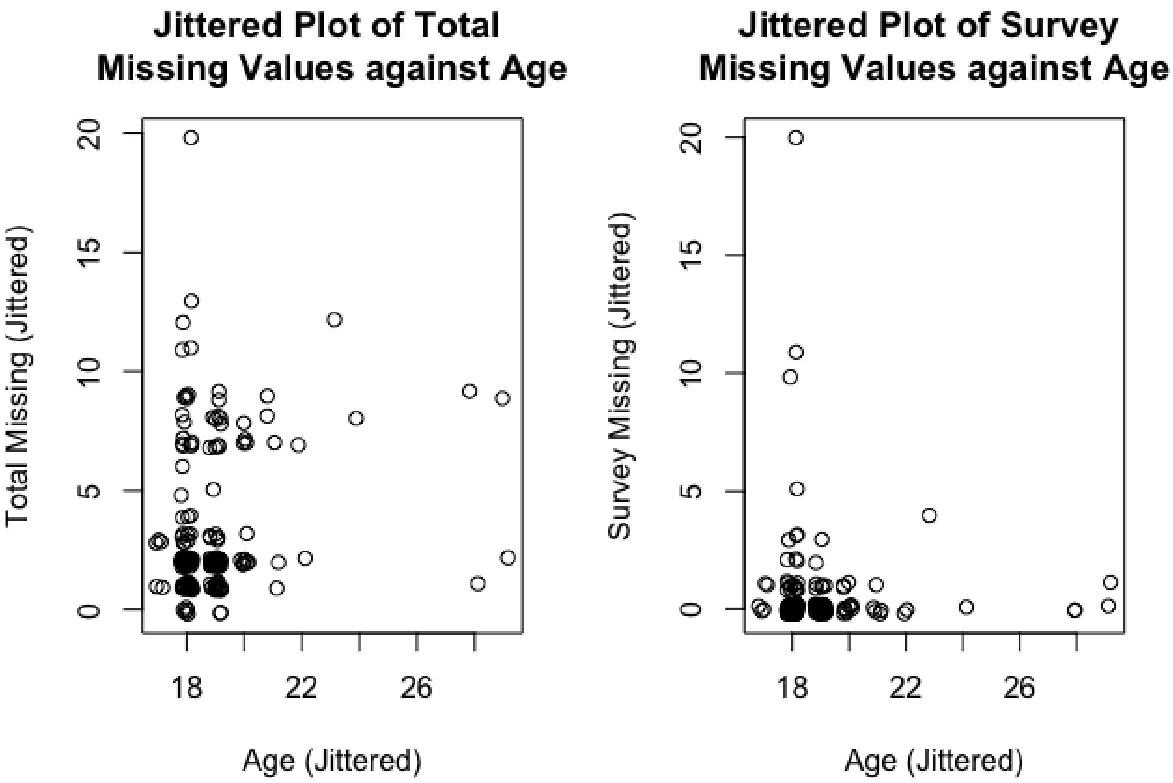

Figure 4: Number of Missing Values against Age

Note that Jittering (the addition of small noise to both Age and the number of Missing Values) was applied to the plots for visibility purposes.
Even though the correlation coefficient for the Total Missing Values appears to be significant, the plot does not show an obvious effect for Age.
For the survey items only, there is clearly no significant relationship between Age and Missing Values.

- A similar analysis was done for the relationship between Past Math Performance and the number of Missing Values.
Here, we found no correlation at all: $r \approx -0.13, p >\approx 0.06$ and $r \approx -0.04, p \approx 0.55$ for Total Missing Values and Survey Missing Values respectively.

- We also tested whether there was a difference between students with different levels of highest education in the number of missing values. The frequencies were compared, similar to what was done before for Gender. Fisher's Exact Test showed no significant difference, with $p \approx 0.86$ and $p \approx 0.84$ for Total Missing and Survey Missing respectively.

- Finally, we tested whether the country of highest education was a predictor in the number of missing values. Here, only the NL/non-NL comparison was made, as there are too many countries to consider otherwise. Fisher's Exact Test showed a significant difference for the total number of missing values ($p \approx 0.04$). However, the difference in means turned out to be very small: NL students have on average 3.25 total missing values and Non-NL students have on average 3.33 average total missing values.
When we look only at the survey items, there was no difference between NL and Non-NL students with $p \approx 0.86$.
This implies that students with their highest education outside of the Netherlands tend to skip some demographic questions more than students from the Netherlands, but for the survey items the missing

values do appear to be MAR, so this should not be a problem for the model.

We conclude that the missing values are Missing At Random and therefore, the second assumption of SEM is satisfied after cleaning the data in the way that we did.
Figures 5 and 6 show the distribution of the total number of missing values and the number of missing values for the survey items only, respectively:

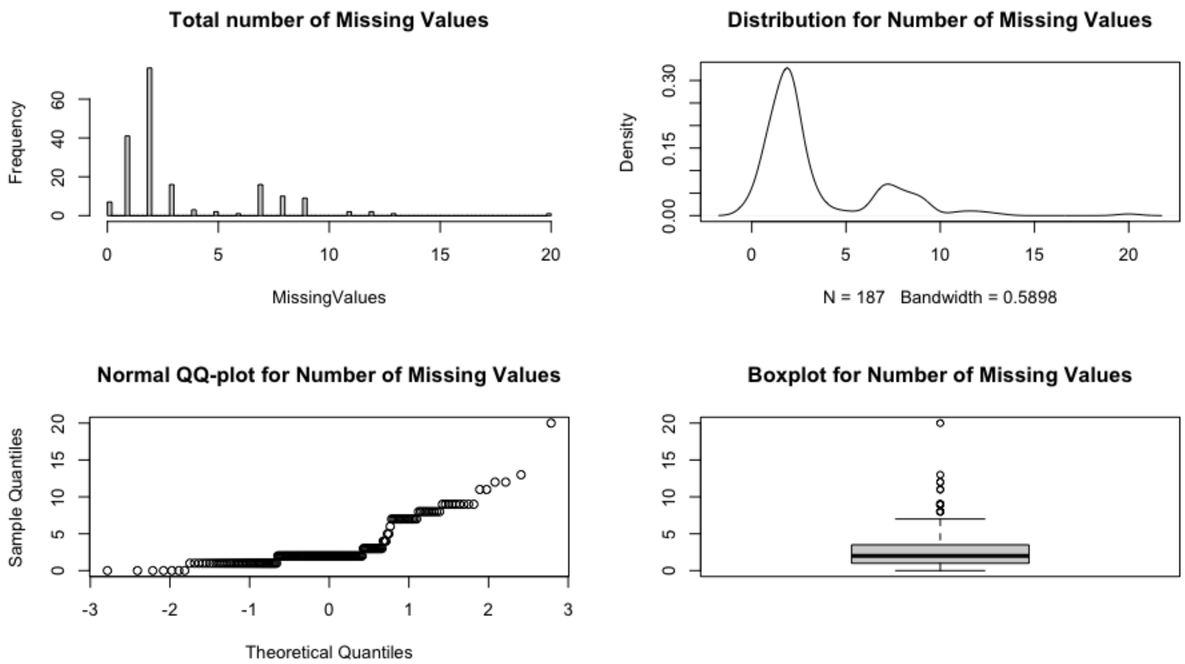

Figure 5: Distribution for the Total number of Missing Values

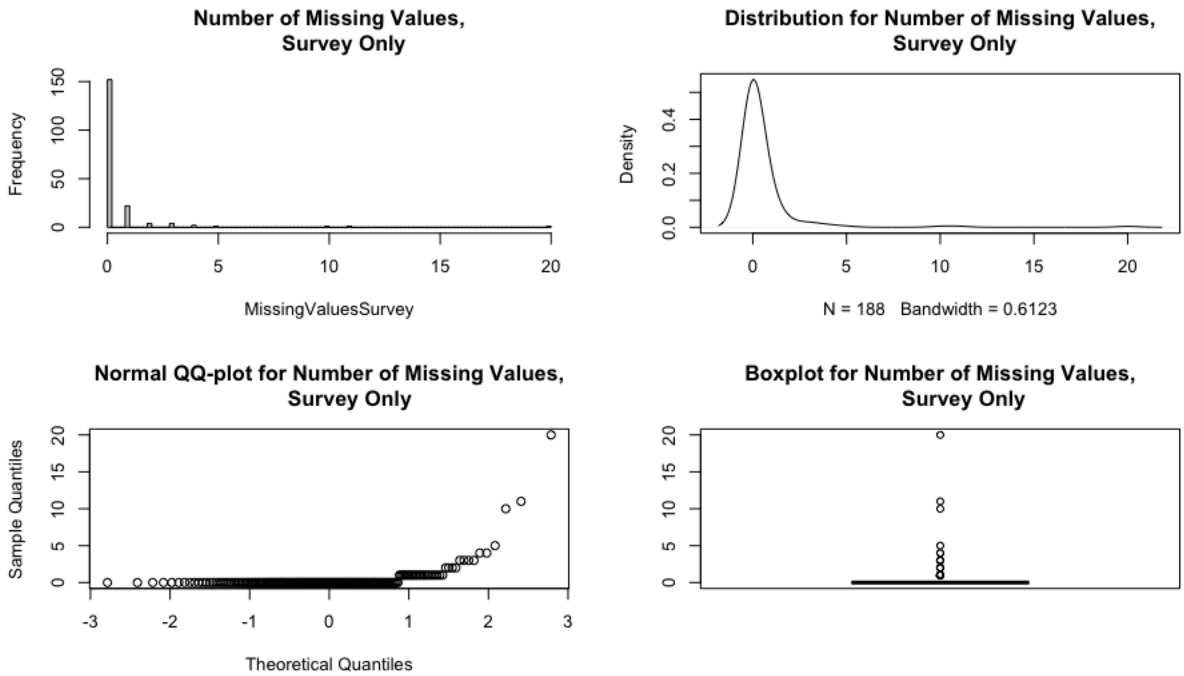

Figure 6: Distribution for the Total number of Missing Values

We see that there appear to be some outliers. For example, there is still one student with 20 missing values, 1 with 12 missings, and 1 with 11. However, we decided not to remove these students, since the missing values do appear to be evenly distributed over all constructs, and therefore can be considered missing at random.

6 Demographics

The data used in this research was collected by PRIME at a Calculus course for the Bachelor 'Computer Science and Engineering' at TU Delft.

After cleaning the data, the total data set consists of 187 students.

Gender: In total, 181 students identified their gender to be either 'male', 'female' or 'other.' Out of these, there were 152 males, 24 females, and 5 others. The 6 missing values were also set to 'other', resulting in a total of 11 others.

Age: 187 students reported their age (1 missing value). The minimum age was 17 and the maximum age was 29. Means and standard deviations are found in table 5

	Overall	Males	Females	Other
Mean	18.9	18.8	18.9	18.9
SD	1.7	1.8	1.4	1.4

Table 5: Means and standard deviations for age

Education Level: 186 students reported their education level (2 missing values). 168 participants' highest education level was secondary education, followed by 5 who completed higher vocational education (HBO), and 12 who entered 'Other'.

Country of highest education: All 187 students reported the country in which they completed their highest level of education. There were 36 different number of countries reported. Most of the students came from the Netherlands ($n = 74$), followed by Romania ($n = 17$), and Belgium and Bulgaria (both $n = 8$).

Also, a new column was created in the data set, transforming the country of highest education to NL/non-NL. As stated before, 74 students reported the Netherlands as the country of their highest education. The remaining 113 students reported a country other than Netherlands.

Past Math Performance: 185 students reported the final grade from their highest level of education for mathematics (2 missing), with 10 being the highest possible grade and 1 being the lowest possible grade. Students' reported score was used as an indicator of their past math performance. Table 6 shows the distribution of the grades students achieved:

Grade	10	$[9, 10)$	$[8, 9)$	$[7, 8)$	$[6, 7)$	$[0, 6)$
Students	34	47	50	37	12	6

Table 6: Past math performance

Note: Past Math Performance needed to be transformed into numerical values for some students (A Was changed to 10, B was changed to 7.5, 6/7 was changed to 6.5, 7/8 to 7.5, etc. Unclear entries were deleted and thus set to missing.

Self-efficacy score: 80 students reported their level of confidence in scoring an 8 or higher in the final exam of this course (107 missing), with 10 being highly certain and 0 being not certain at all. This score was used as an indicator of self-efficacy. Table 7

Self-Efficacy Score	10	9	8	7	6	5	4	3	2	1
Students	0	0	0	0	0	33	17	12	13	5

Table 7: Self-Efficacy Scores

7 Results

Before we take a look at the results with respect to the correlations obtained by the model, we need to make sure the that the $t-$rule holds for each model, and also that the goodness-of-fit measures are in the right ranges. For a complete description of the model we refer to Chapter 3, and for an explanation of the $t-$rule and the goodness-of-fit measures, we refer to Chapter 2, Sections 2.3 and 2.5 respectively.

The following Table shows what survey items are initially used to define our latent variables. The content of these items can be found in Chapter 4. We will need the information from Table 8 to determine whether the $t-$rule holds.

Latent variable	# items (p)	Items used
Monitoring	15	SM1, SM3, SM4, SM5, SM6, SM7, SM8, SM9, SM10, SM11, SM12, SM13, SM14, SM15, SM16
Scaffolding	14	SM17, SM18, SM19, SM20, SM21, SM22, SM23, SM24, SM25, SM26, SM27, SM28, SM29, SM30
Autonomy Satisfation	4	Needs1, Needs7, Needs13, Needs19
Autonomy Frustration	4	Needs2, Needs8, Needs14, Needs20
Relatedness Satisfaction	4	Needs3, Needs9, Needs15, Needs21
Relatedness Frustration	4	Needs4, Needs10, Needs16, Needs22
Competence Satisfaction	4	Needs5, Needs11, Needs17, Needs23
Competence Frustration	4	Needs6, Needs12, Needs18, Needs24
Autonomous Motivation	8	M2, M4, M7, M8, M11, M13, M15, M16
Controlled Motivation	8	M1, M3, M5, M6, M9, M10, M12, M14
Academic Performance	1	Final Result

Table 8: Survey items initially used to define latent variables

In order to have a better understanding of the nature of the relationships between the latent variables, the complete model was split up into three parts. The first part concerns Monitoring and Scaffolding, and their relationships with four Needs: Relatedness and Competence, both Satisfaction and Frustration. The second part is about the relationships between all six Needs and the two types of Motivation. The third part relates the two types of Motivation to Academic Performance. Finally, the complete model is shown, combining all three parts of the model. Note that in the complete model, we expect different results for the relationships, as the complete model includes many mediation effects, as explained in Chapter 2, Section 2.2.2.

In the next Section, all models will be analyzed separately.

7.1 Results for the initial models

7.1.1 Part 1: Monitoring, Scaffolding and Needs

The first part of the model is about the perceived Monitoring and Scaffolding cues, and the relationships with the Needs. This is presented in the following graph:

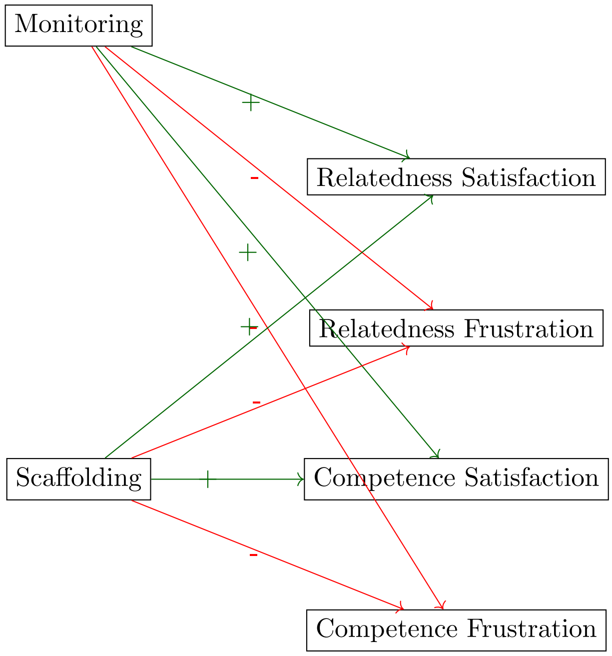

Figure 7: Graphical Presentation of the first part of the hypothesized model

We start with checking the minimal condition of identifiability, which is the $t-$rule as stated in equation 7 explained in Chapter 2.3. From Table 8 and Figure 7 we deduce that $p = 45$ and $z = 8$ and therefore we obtain

$$\frac{1}{2}(45)(46) = 1035 \geq 98 = 2(45) + 8 \tag{18}$$

and so the $t-$rule holds.

After implementation, the model was tested, which resulted in the following goodness-of-fit measures:

Goodness-of-fit measure	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
	0	0.879	0.871	(0.069,0.079)	0.109

Table 9: Goodness-of-fit measures for the original model considering Monitoring, Scaffolding and the six Needs

The $p-$value for the χ^2-statistic is 0, but this was to be expected, as explained in Chapter 2.5. However, the other goodness-of-fit measures are also not good enough. The CFI, TLI and RMSEA are not acceptable, and the SRMR can even be considered bad (see Table 1 in Chapter 2.5).
Adjustments to this part of the model need to be made in order to make the model fit the data better by making use of the modification indices and, if necessary, by removing some items from the survey to improve the SRMR.
This is done in Chapter 7.2.

7.1.2 Part 2: Needs and Motivation

The second part of the model is about the Needs and their relationships with Autonomous and Controlled Motivation. This is presented in the following graph:

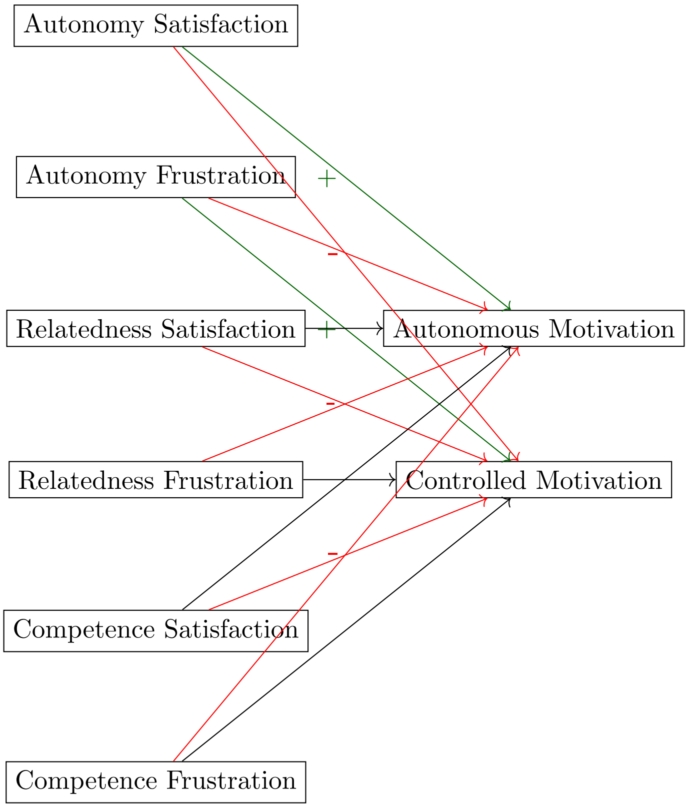

Figure 8: Graphical Presentation of the second part of the hypothesized model

From Table 8 and Figure 8 we deduce that $p = 40$ and $z = 12$ and therefore we obtain

$$\frac{1}{2}(40)(41) = 820 \geq 92 = 2(40) + 12 \tag{19}$$

and so the $t-$rule holds.

The model was implemented and tested, resulting in the following goodness-of-fit measures:

Goodness-of-fit measure	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
	0	0.934	0.928	(0.048,0.061)	0.079

Table 10: Goodness-of-fit measures for the original model considering the six Needs and Motivation

As expected, the $p-$value for the χ^2-statistic is again 0. However, we see that the TLI and the CFI are acceptable, and so is the SRMR (see Table 1). The RMSEA is acceptable to good. We expect some small adjustments need to be made in order for all measures to be acceptable to good. This is done in Chapter 7.2.

7.1.3 Part 3: Motivation and Academic Performance

The third part of the model is about Autonomous and Controlled Motivation and their relationships with Academic Performance. This is presented in the following graph:

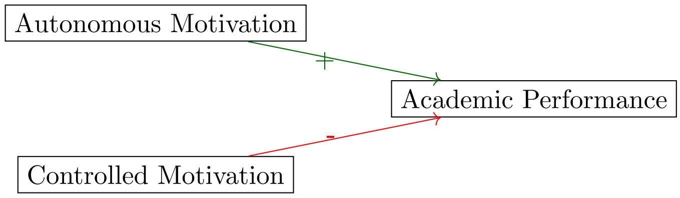

Figure 9: Graphical presentation of the third part of the hypothesized model

From Table 8 and Figure 9 we deduce that $p = 17$ and $z = 2$ and therefore we obtain

$$\frac{1}{2}(17)(18) = 136 \geq 36 = 2(17) + 2 \tag{20}$$

and so the $t-$rule holds.

After implementation, the following results were obtained for the goodness-of-fit measures:

Goodness-of-fit measure	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
	0	0.913	0.899	(0.097,0.122)	0.1

Table 11: Goodness-of-fit measures for the original model considering Motivation and Academic Performance

As expected, the $p-$value for the χ^2-statistic is again 0. However, we see that the CFI is acceptable but the TLI is not. The RMSEA and SRMR are even considered bad. Adjustments are necessary to improve the goodness-of-fit.
This is described in Chapter 7.2.

7.1.4 The complete model

Finally, the complete model is represented as already shown in Chapter 3.3:

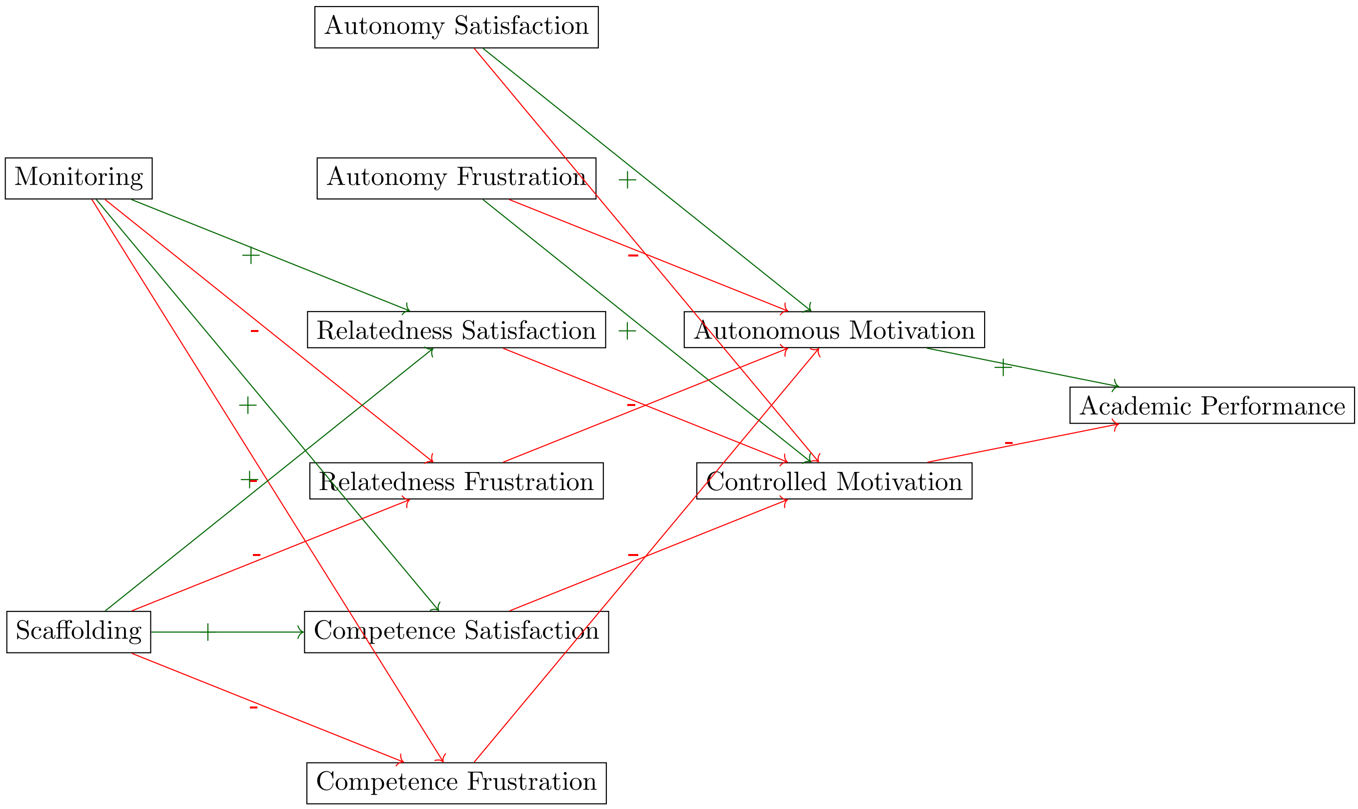

Figure 10: Graphical Presentation of the hypothesized model

Again, we need to check if the $t-$rule holds. From Table 8 and Figure 10 we deduce that $p = 70$ and $z = 20$ and therefore we obtain

$$\frac{1}{2}(70)(71) = 2485 \geq 160 = 2(70) + 20 \tag{21}$$

and so the $t-$rule holds.

The initial complete model resulted in the following goodness-of-fit measures:

Goodness-of-fit measure	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
	0	0.644	0.640	(0.084,0.09)	0.103

Table 12: Goodness-of-fit measures for the original model considering the complete model

As expected, the $p-$value for the χ^2-statistic is again 0. We see that the CFI and TLI are not even close to acceptable, and the RMSEA and SRMR are far too high (see Table 1). Adjustments need to be made to the separate models described above, and then combined to form an improved complete model. This is described in Chapter 7.2.

7.2 Improving the models step by step

In the next Sections, all parts of the model are improved based on Modification Indices, as explained in Chapter 2, Section 2.6.

7.2.1 Part 1: Monitoring, Scaffolding and Needs

As we noted in the previous Chapter, adjustments need to be made to this part of the model. We use `R` to obtain modification indices for individual regressions. As explained in Chapter 2.5, the model will have a poorer fit if the measured variables are heavily correlated to one another. If these regressions are included in the model, and are therefore taken into account, the goodness-of-fit will improve. A regression with the highest modification index will produce the best improvement in the goodness-of-fit, so we start by adding those regressions that give us the highest modification index (we will of course manually check whether it makes sense to add these regressions based on the content of the items). We will proceed with this step wise improving until we see that either the goodness-of-fit is sufficient, or the goodness-of-fit is no longer improving significantly. Table **??** shows the process of the improvements.

It is important that we do not simply add all suggestions to the model. We need to make sure that the regressions make sense theoretically, as explained in Chapter 2.7.
For example, we obtain a high modification index for the regression for MSC, items 22 and 23. When we look at these items, we see that they are both Scaffolding items:

- 22: The lecturer asks questions in a way I understand.
- 23: The lecturer asks questions that help me gain understanding of the subject matter.

When we compare these two items, they are different questions but are similar enough to expect a correlation between the two. Therefore, we allow this correlation to be a part of our model. Similarly, the other suggested improvements to the model were checked and added.

	# parameters	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
Initial model	239	0	0.879	0.871	(0.069,0.079)	0.109
Adding regressions (1) MSC: 22 ∼∼ 23 28 ∼∼ 29 29 ∼∼ 30 9 ∼∼ 10 16 ∼∼ 18	244	0	0.897	0.889	(0.064,0.074)	0.106
Adding regressions (2): MSC: 28 ∼∼ 30 25 ∼∼ 27 19 ∼∼ 20 14 ∼∼ 27 Needs: 15 ∼∼ 21	249	0	0.904	0.897	(0.061,0.071)	0.103
Adding regressions (3): MSC: 20 ∼∼ 23 5 ∼∼ 27 5 ∼∼ 24 24 ∼∼ 27 5 ∼∼ 14	254	0	0.911	0.903	(0.059,0.069)	0.100

Table 13: Goodness-of-fit measures for the original model and each improvement considering Monitoring, Scaffolding and the six Needs (first part of improvements)

Note that the Table resumes on the next page.

	# parameters	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
Adding regressions (4): MSC: 21 $\sim\sim$ 24 19 $\sim\sim$ 23 23 $\sim\sim$ 24 14 $\sim\sim$ 18 12 $\sim\sim$ 18	259	0	0.916	0.909	(0.057,0.068)	0.098
Adding regressions (5): MSC: 24 $\sim\sim$ 25 5 $\sim\sim$ 20 13 $\sim\sim$ 23 14 $\sim\sim$ 22 5 $\sim\sim$ 9	264	0	0.921	0.914	(0.55,0.66)	0.096
Adding regressions (6): MSC: 22 $\sim\sim$ 25 5 $\sim\sim$ 19 5 $\sim\sim$ 23 14 $\sim\sim$ 25 18 $\sim\sim$ 22	269	0	0.925	0.918	(0.054,0.065)	0.094

Table 14: Goodness-of-fit measures for the original model and each improvement considering Monitoring, Scaffolding and the six Needs (second part of improvements)

From Tables 13 and 14 we conclude that the CFI is good enough with improvement (2) and the TLI with improvement (3) (see Table 1 for an interpretation of the values). However, the RMSEA and SRMR are not improving enough, even at improvement (6) they are still not good enough. Therefore, we move on to another method of improving the model, based on local fit measures instead of global fit measures. Some items may need to be removed as they are too inconsistent with the other items within its construct. We compute the mean residual covariance per item over all items and remove the items with the highest absolute mean residuals until we obtain acceptable RMSEA and SRMR:

	# parameters	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
Previous model	269	0	0.925	0.918	(0.054,0.065)	0.094
Removing SM14 (abs.res=2.29)	259	0	0.927	0.920	(0.053,0.064)	0.093
Removing SM27 (res 2.01)	252	0	0.932	0.925	(0.052,0.063)	0.089
Removing SM18 (res 1.92)	244	0	0.938	0.932	(0.050,0.061)	0.087
Removing SM26 (res 1.84)	239	0	0.948	0.942	(0.045,0.058)	0.083
Removing SM3 (res 1.77)	229	0	0.952	0.946	(0.044,0.057)	0.081
Removing SM28 (res 1.77)	222	0	0.955	0.950	(0.043,0.057)	0.079

Table 15: Goodness-of-fit measures for each improvement after removing individual items considering Monitoring, Scaffolding and the six Needs

We note that after the removal of six individual items concerning Monitoring and Scaffolding , we obtain acceptable results for all four measures of interest. The CFI and TLI can even be considered good now (see Table 1).

Since we removed some items, we need to check whether the $t-$rule still holds. Equation 18 no longer applies here. We now have twelve items for Monitoring and ten items for Scaffolding instead of fifteen and fourteen respectively. The four Needs are all still based on four items. This means that we now have

$$\frac{1}{2}(38)(39) = 546 \geq 84 = 2(38) + 8 \tag{22}$$

and so the $t-$rule still holds.

The following graph shows the final model for this first part:

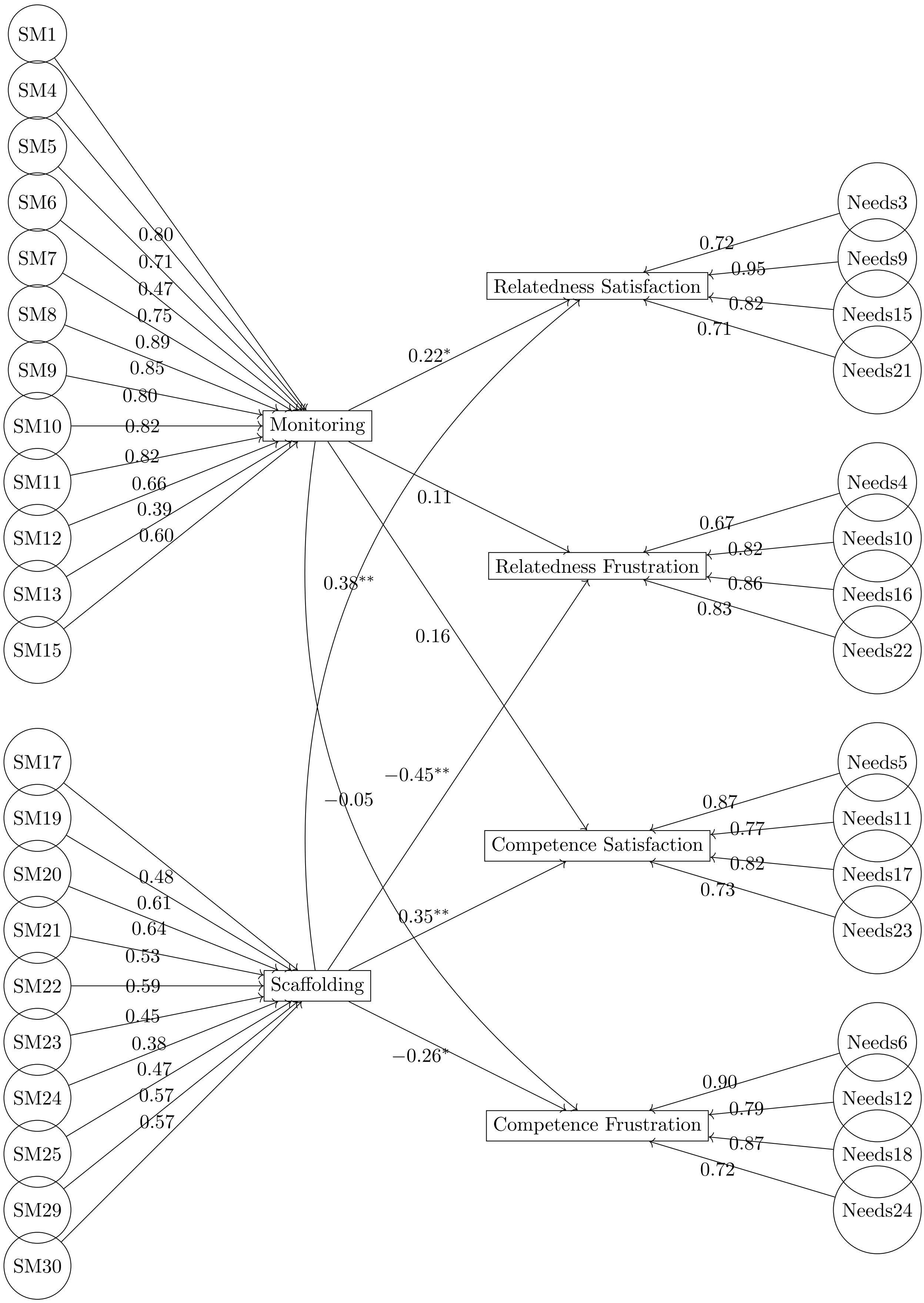

Figure 11: Graphical Presentation of the first part of the final model

The values indicate standardized estimates for the factor loadings from the items to the latent variables, and standardized estimates for the regression coefficients between latent variables.
Statistically significant relationships are denoted with * and ** at levels $\alpha = 0.05$ and $\alpha = 0.01$ respectively.

7.2.2 Part 2: Needs and Motivation

This part of the model was initially almost good enough already. Therefore, we add only two regressions to the model:

	# parameters	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
Initial model	224	0	0.934	0.928	(0.048,0.06)	0.08
Adding regressions:						
Motivation $1 \sim\sim 14$ $5 \sim\sim 12$	226	0	0.949	0.944	(0.041,0.054)	0.076

Table 16: Goodness-of-fit measures for each improvement considering the six Needs and Motivation

We note that these small improvements are indeed sufficient. The RMSEA is now acceptable to good. The CFI, TLI and SRMR are acceptable.
Since we did not remove any items, we know that the $t-$rule from equation 19 still holds as the equation remains unaltered.

The following graph shows the final model for this second part:

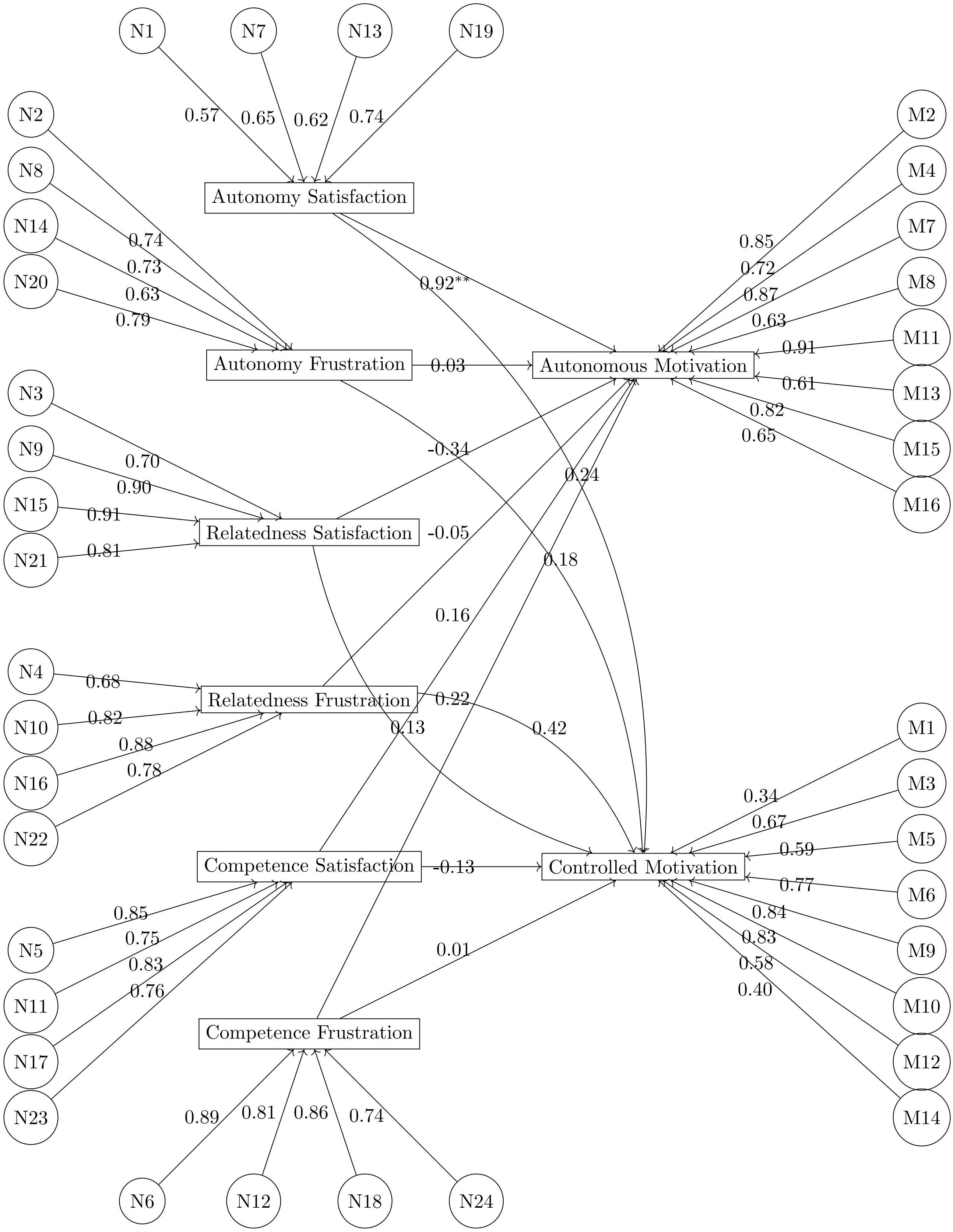

Figure 12: Graphical Presentation of the second part of the final model

Note that the twenty-four Needs are this time indicated by N1 up to N24 for readability purposes.

The values indicate standardized estimates for the factor loadings from the items to the latent variables, and standardized estimates for the regression coefficients between latent variables.
Again, statistically significant relationships are denoted with * and ** at levels $\alpha = 0.05$ and $\alpha = 0.01$ respectively.
The reported estimates can also be found later in Table 21, as it may be unclear in this Figure what estimate belongs to what arrow.

7.2.3 Part 3: Motivation and Academic Performance

Similar to the first two parts of the model, we try to improve the third part of the model by manually adding regressions based on the modification indices.

	# parameters	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
Initial model	85	0	0.913	0.899	(0.097,0.122)	0.1
Adding regressions (1): Motivation 1 ~~ 14 5 ~~ 12 6 ~~ 9 6 ~~ 12 13 ~~ 16	90	0	0.969	0.963	(0.052,0.081)	0.073
Adding regressions (2): Motivation 8 ~~ 13 8 ~~ 16	92	0	0.982	0.978	(0.034,0.067)	0.067

Table 17: Goodness-of-fit measures for each improvement considering Motivation and Academic Performance

Table 17 shows that the CFI and TLI are good and the SRMR is acceptable, but the RMSEA is not. However, the remaining Modification Indices are too small (¡10.83, for an explanation of this, we refer to Section 2.7), so there is no use in adding more regressions.
Since the SRMR is sufficient, it does not make sense to remove items either.
However, it is possible to adjust our variable Academic Performance.
It is currently defined to be the final result of a student for the course as submitted into Osiris.
However, other possibilities are to convert this result into an ordinal or even binary variable. After all, the other measured variables are all ordinal in nature. Also, a distinction can be made in how to handle the missing values. It is possible to omit the students without a final grade, but we could also set their grade to 0.

Based on the previous model, several options on how to handle Academic Performance are taken into account. The results are shown in Table 18:

	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
Scale unaltered NA is omitted	0	0.982	0.978	(0.034,0.067)	0.067
Scale unaltered NA is set to 0	0	0.983	0.979	(0.033,0.066)	0.065
Scale 2 quantiles (≥ 6 Pass/< 6 Fail) NA is omitted	0	0.989	0.986	(0.018,0.058)	0.081
Scale 2 quantiles (≥ 6 Pass/< 6 Fail) NA is set to 0	0	0.993	0.991	(0, 0.052)	0.066
Scale 4 quantiles: (0-2.5, 2.6-5, 5.1-7.5, 7.6-10) NA is omitted	0	0.992	0.990	(0, 0.053)	0.067
Scale 4 quantiles: (0-2.5, 2.6-5, 5.1-7.5, 7.6-10) NA is set to 0	0.063	0.992	0.990	(0, 0.053)	0.066
Scale 10 quantiles: round up to integer 1-10 NA is omitted	0.045	0.991	0.989	(0.006, 0.054)	0.066
Scale 10 quantiles: round up to integer 1-10 NA is set to 0	0.060	0.992	0.990	(0, 0.053)	0.066

Table 18: Goodness-of-fit measures for several ways to define Academic Performance

Table 18 shows that the last five options all result in sufficiently good fits. However, the best fit is obtained when treating Academic Performance as a 4-quantile scale, with NA set to 0 (and therefore set to 1 due to the scaling). We see here that even $p(\chi^2) > 0.05$. We now have a good CFI and TLI, an acceptable to good RMSEA and an acceptable SRMR (see Table 1). Therefore, students performance is set to 1 if they have a grade lower than 2.5 or no grade at all; it is set to 2 if they score between 2.6 and 5; it is set to 3 if they score between 5.1 and 7.5 and finally their grades are set to 4 if they score higher than 7.5.

Since we did not remove any items, we know that the $t-$rule from equation 20 still holds as the equation remains unaltered.

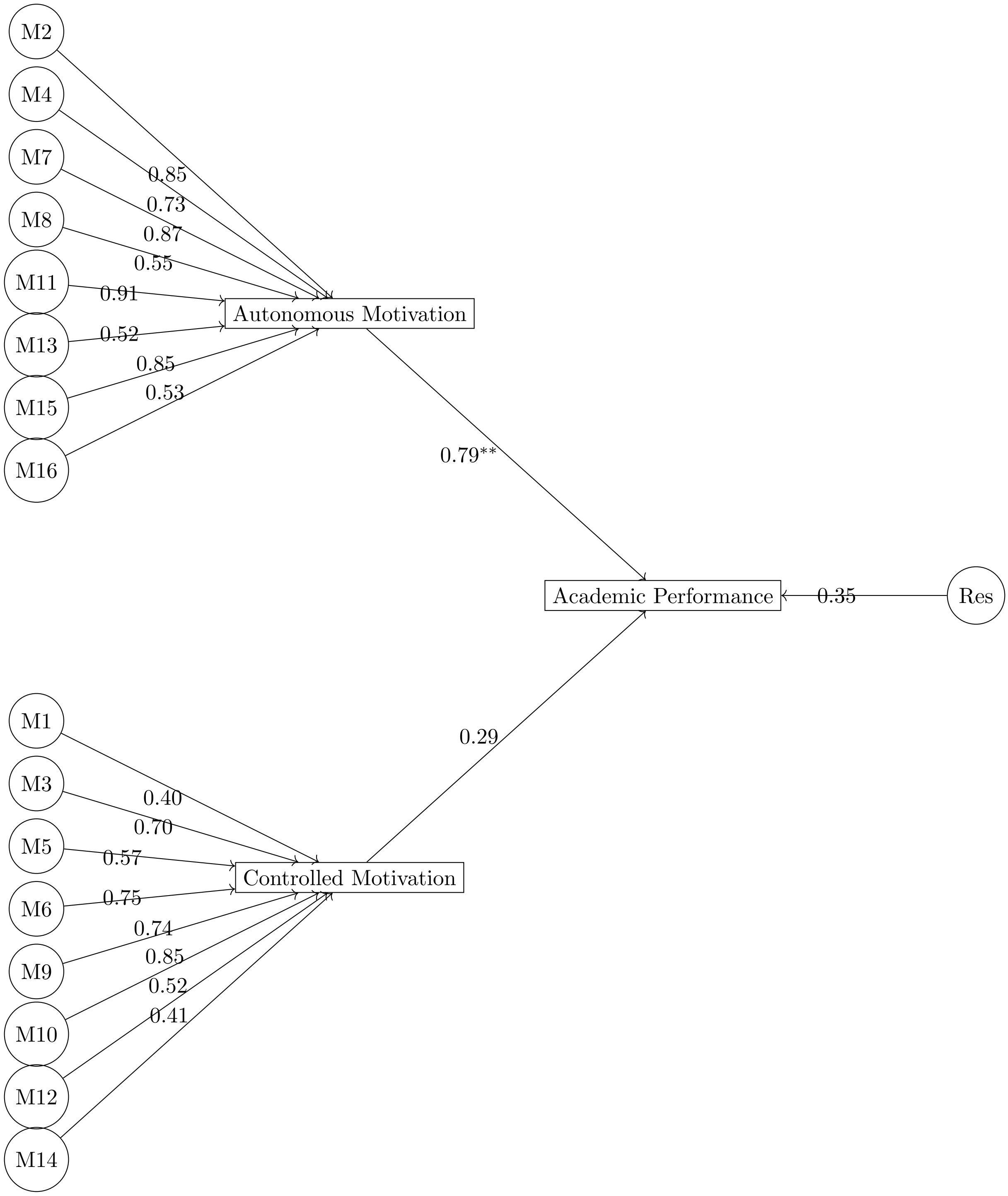

Figure 13: Graphical Presentation of the third part of the final model

Note that the item 'Res' is the final result for the course split into four quantiles, as explained above. The values indicate standardized estimates for the factor loadings from the items to the latent variables, and standardized estimates for the regression coefficients between latent variables.

To summarize, we have the following goodness-of-fit measures for the three submodels:

Goodness-of-fit measure	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
1: Needs $\sim\sim$ Mon/Scaf	0	0.955	0.950	(0.043,0.057)	0.079
2: Motivation $\sim\sim$ Needs	0	0.949	0.944	(0.041,0.054)	0.076
3: AP $\sim\sim$ Motivation	0.063	0.992	0.990	(0,0.053)	0.066

Table 19: Improved goodness-of-fit measures for all individual parts of the model

We are very content with the results, as all goodness-of-fit measures are either acceptable to good or good.

The factor loadings and regression coefficients can be read from the graphical presentations in Figures 11, 12 and 13. The complete outputs from `R` can be found in the appendix.

7.3 Results for the final model

Similar to the first part of the model, we need to check whether the complete model is identified by checking if the $t-$rule still holds after some items were removed. Equation 21 no longer applies. We now have

$$\frac{1}{2}(63)(64) = 2016 \geq 146 = 2(63) + 20 \tag{23}$$

and so the $t-$rule still holds.

Based on the improvements for the individual parts, as described in the previous subsections, we have come to the following goodness-of-fit measures for the complete model:

Goodness-of-fit measure	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
	0	0.974	0.973	(0.053,0.061)	0.086

Table 20: Improved goodness-of-fit measures for the complete model

We notice that we have obtained very good results for the CFI and the TLI. The RMSEA can be considered to be acceptable, as it lies with 90% certainty between 0.053 and 0.061. However, the SRMR is too large, as it exceeds 0.08.
The results are much better than they were before, but still not good enough. We conclude that the individual parts of the model are more reliable to draw conclusions from. However, we will still present the result for our final model:

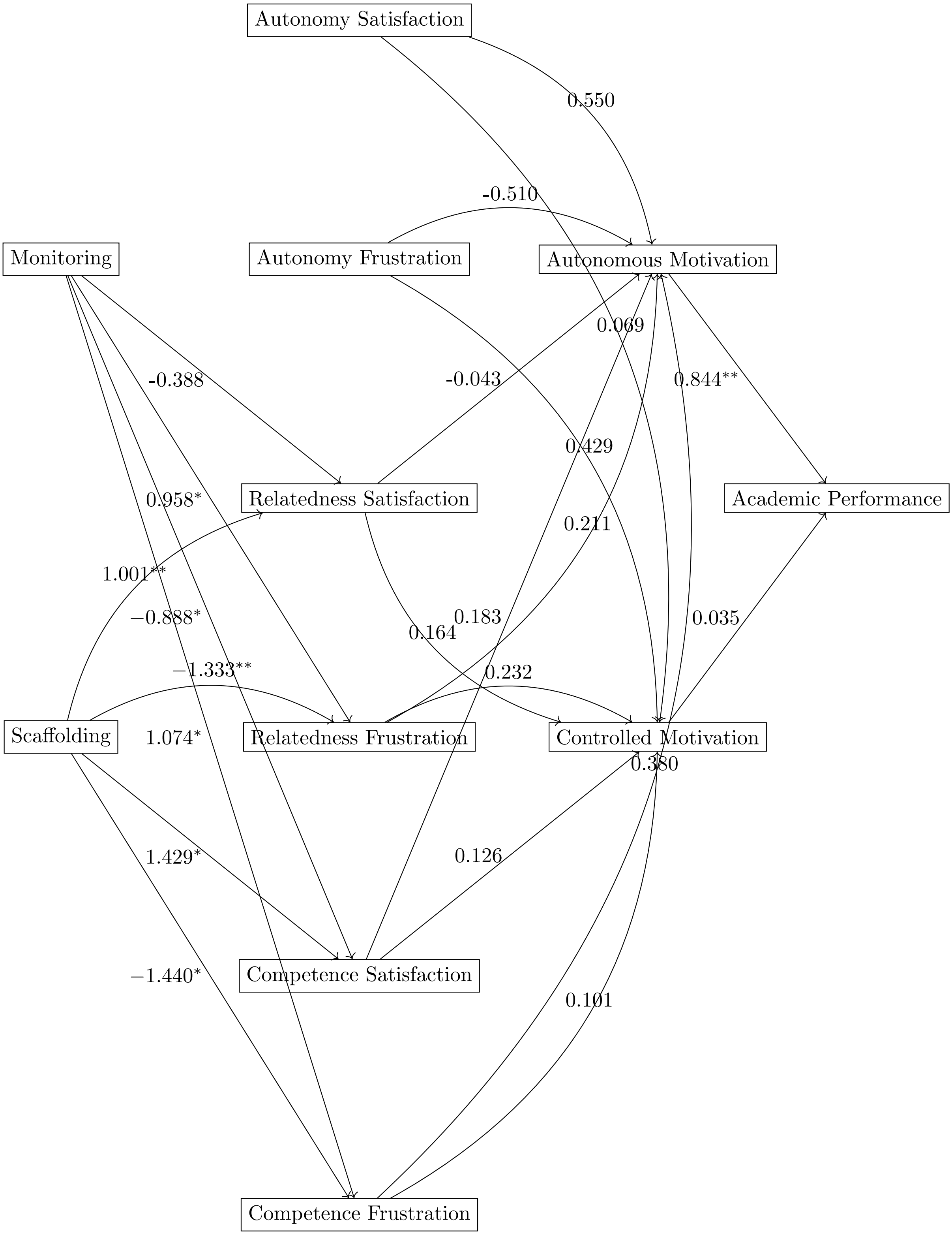

Figure 14: Graphical Presentation of the combined model

Graph 14 can also be presented with the following Table:

	Mon	Scaf	AS	AF	RS	RF	CS	CF	AM	CM	AP
Mon											
Scaf											
AS											
AF											
RS	−0.39	1.00*									
RF	0.96*	−1.33*									
CS	−0.89*	1.43*									
CF	1.07*	−1.44*									
AM			0.55	-0.51	-0.04	0.21	0.18	0.38			
CM			0.07	0.43	0.16	0.23	0.13	0.10			
AP									0.84*	0.04	

Table 21: Standardized regression coefficients for the total model

	Mon	Scaf	AS	AF	RS	RF	CS	CF	AM	CM	AP
Mon											
Scaf											
AS											
AF											
RS	0.22*	0.38*									
RF	0.11	−0.45*									
CS	0.16	0.35*									
CF	−0.05	−0.26*									
AM			0.92*	0.03	-0.34	-0.05	0.16	0.22			
CM			0.24	0.18	0.13	0.42	-0.13	0.01			
AP									0.79*	0.29	

Table 22: Standardized regression coefficients for all three parts of the model combined

Compared to the individual parts, which is presented in Table 22, interesting observations can be made. First of all, we see completely different numbers. This is not very strange, considering the much higher number of levels we have in the total model. The total model covers all mediation effects, which are non-existent in the separate models. For example, for the relationship between Autonomous Motivation and Relatedness Satisfaction, we obtain $\beta = -0.04$ in the total model, but we obtain $\beta = -0.34$ in the second sub-model, considering only the six Needs and Motivation. These numbers differ greatly, because in the first case, the effect of Relatedness Satisfaction on Autonomous Motivation was adjusted for the effect of the teacher cues Monitoring and Scaffolding.
This mediation effect will become more apparent in the next subsection where we discuss the hypotheses, as some of the hypotheses involve mediation effects specifically.

Overall, we do see some interesting similarities, though. We see that the relationship between Scaffolding and the four Needs RS, RF, CS and CF are similar (despite the numbers, the signs and significance remain the same). The relationships between the Needs and Motivation are small or non-existent in both cases (with the exception of Autonomy Satisfaction, having a significant effect on Autonomous Motivation in the sub-model). Finally, we see that Autonomous Motivation has a lot of influence on Academic Performance in both cases, where Controlled Motivation seems to be irrelevant.

We must note though that we need to be careful with our interpretations, considering the goodness-of-fit for the total model.

7.4 Testing the hypotheses

Based on the results from Chapter 7.2 we aim to answer the following hypotheses formulated in Chapter 3.2:

1. Perceived teacher Scaffolding and Monitoring cues are positively related to Competence and Relatedness Satisfaction and negatively related to Competence and Relatedness Frustration
2. Autonomy Satisfaction is positively related to Autonomous Motivation and negatively to Controlled Motivation
3. Autonomy Frustration is positively related to Controlled Motivation and negatively to Autonomous Motivation
4. Competence and Relatedness Satisfaction are negatively related to Controlled Motivation
5. Competence and Relatedness Frustration are negatively related to Autonomous Motivation, but less strongly than the relationships between Competence and Relatedness Satisfaction and Autonomous Motivation, and Competence and Relatedness Frustration and Controlled Motivation
6. Competence and Relatedness Satisfaction positively mediate the relationship of perceived teacher cues and Autonomous Motivation
7. Competence and Relatedness Frustration negatively mediate the relationship of perceived teacher cues and Controlled Motivation
8. There is a positive relationship between Autonomous Motivation and performance and a negative, but less strong, relationship between Controlled Motivation and performance

We consult Table 21 and the complete `R` output (which can be found in the appendix) in order to test these hypotheses.

Note that hypotheses 6 and 7 involve mediation effects, which was explained in Chapter 2.2.2.

Hypothesis 1: Perceived teacher Scaffolding and Monitoring cues are positively related to Competence and Relatedness Satisfaction and negatively related to Competence and Relatedness Frustration

This hypothesis can be tested directly based on the results we have from the first part of the model, as shown in Figure 11. To discuss the results, this hypothesis is split into eight distinct hypotheses:

We expect

1. A positive relationship between Monitoring cues and Competence Satisfaction

   Whereas this may appear to be true, as we obtain a positive coefficient, it is not statistically significant ($\beta \approx 0.16, p > 0.05$) so we conclude it is not reliable to predict Competence Satisfaction based on Monitoring cues.

2. A positive relationship between Scaffolding cues and Competence Satisfaction

   This is supported by our results: $\beta \approx 0.35, p \approx 0.003$.

3. A positive relationship between Monitoring cues and Relatedness Satisfaction

   This relationship is also obtained in our model: we find $\beta \approx 0.22, p \approx 0.014$.

4. A positive relationship between Scaffolding cues and Relatedness Satisfaction

   Again, a relationship that is confirmed by our model: $\beta \approx 0.38, p \approx 0.001$.

5. A negative relationship between Monitoring cues and Competence Frustration

   Even though we obtain a negative regression coefficient, the relationship obtained is not statistically significant: $\beta \approx -0.05, p > 0.05$. We therefore conclude that is not reliable to predict Competence Frustrationbased on Monitoring cues.

6. A negative relationship between Scaffolding cues and Competence Frustration

   This relationship was indeed obtained by our model: $\beta \approx -0.26, p \approx 0.038$.

7. A negative relationship between Monitoring cues and Relatedness Frustration

   Based on our model, no relationship was found between Relatedness Frustration and Monitoring cues: $\beta \approx 0.11, p > 0.05$. Also, though there is no significance, the relationship tends to be positive rather than negative.

8. A negative relationship between Scaffolding cues and Relatedness Frustration

   This final part of our hypothesis is supported by our model: $\beta \approx 0.45, p \approx 0.001$.

Based on our results we can therefore confirm five out of the eight sub-hypotheses. Interestingly, all four sub-hypotheses with respect to Scaffolding are confirmed, whereas only one out of the four sub-hypotheses concerning Monitoring is confirmed.

Hypothesis 2: Autonomy Satisfaction is positively related to Autonomous Motivation and negatively to Controlled Motivation

The second model concerning the six Needs and two types of Motivation, give us standardized regression coefficients based on all six Needs. We are therefore dealing with a situation of multicollinearity as the six Needs are not uncorrelated, which complicates how we should interpret the regression coefficients. The regression coefficient between Autonomy Satisfaction and Autonomous Motivation may be different when the other five Needs are not taken into account.
Based on the second model though, we can say that the first part of this hypothesis is confirmed by our results: Autonomy Satisfaction appears to positively influence Autonomous Motivation ($\beta \approx 0.92, p \approx 0.002$). However, for Controlled Motivation, we obtain no statistically significant relationship ($\beta \approx 0.24, p > 0.05$).

However, a separate model, looking only at Autonomy Satisfaction and the two types of Motivation, will give us more information to answer this hypothesis correctly. This model is given by Figure 15. Again, we need to check for identification:

$$\frac{1}{2}(20)(21) = 210 \geq 42 = 2(20) + 2 \tag{24}$$

and so the $t-$rule holds.

The following goodness-of-fit measures were obtained:

Goodness-of-fit measure	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
	0	0.959	0.953	(0.054,0.078)	0.076

Table 23: Goodness-of-fit measures for Hypothesis 2

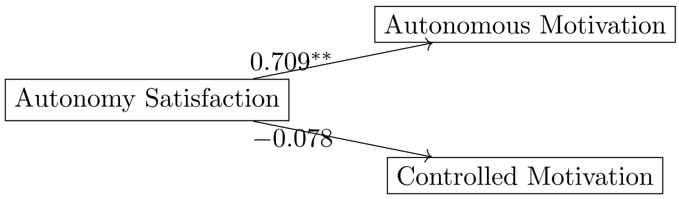

Figure 15: Graphical Presentation of the model for Hypothesis 2

We conclude that Autonomous Motivation is positively influenced by Autonomy Satisfaction($\beta = 0.709, p \approx 0.000$), but Autonomy Satisfaction does not have any significant influence on Controlled Motivation ($\beta = -0.078, p > 0.05$).
Because we have an interval for the RMSEA that is not contained within $(0, 0.06)$, we need to be careful interpreting our results though.

Hypothesis 3: Autonomy Frustration is positively related to Controlled Motivation and negatively to Autonomous Motivation

Based on Figure 12, this hypothesis is not confirmed by our model: we obtain a very small, not statistically significant standardized regression coefficient for the relationship between Autonomous Motivation and Autonomy Frustration($\beta \approx 0.03.p > 0.05$), and another non-significant positive coefficient for Controlled Motivation and Autonomy Frustration($\beta \approx 0.18, p > 0.05$).
However, similar to the previous hypothesis, it is better to consider this need separately instead of together with the five other Needs. The corresponding model is given by Figure 16.
We know the model is identified as the equation for the $t-$rule is exactly the same as the one for the previous hypothesis (equation 24).

The following goodness-of-fit measures were obtained:

Goodness-of-fit measure	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
	0	0.958	0.952	(0.055,0.078)	0.078

Table 24: Goodness-of-fit measures for Hypothesis 3

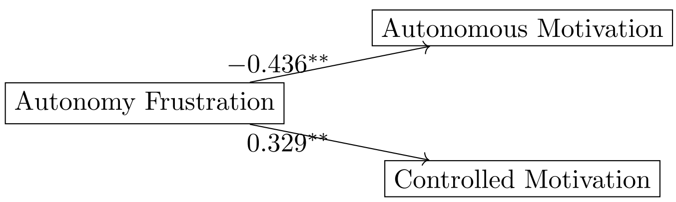

Figure 16: Graphical Presentation of the model for Hypothesis 3

We see that when we consider Autonomy Frustration separately, we do obtain significant results as hypothesized: Autonomous Motivation is negatively influenced by Autonomy Frustration ($\beta = -0.436, p \approx 0.000$), and Controlled Motivation is positively influenced by Autonomy Frustration ($\beta = 0.329, p \approx 0.001$).
However, we must take into account that the RMSEA is out of bounds.

Hypothesis 4: Competence and Relatedness Satisfaction are negatively related to Controlled Motivation

Again, the model in Figure 12 does not seem to support these claims, as the apparent negative relationship for Competence Satisfaction is not statistically significant: $\beta \approx -0.13, p > 0.05$. Also, we do not obtain a negative relationship between Relatedness Satisfaction and Controlled Motivation: ($\beta \approx -0.13, p > 0.05$). Also, when we create a separate model as shown in Figure 17 we see the $t-$rule holds, as we have

$$\frac{1}{2}(16)(17) = 136 \geq 34 = 2(16) + 2 \tag{25}$$

and we obtain the following results:

Goodness-of-fit measure	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
	0	0.979	0.974	(0.042,0.075)	0.071

Table 25: Goodness-of-fit measures for Hypothesis 4

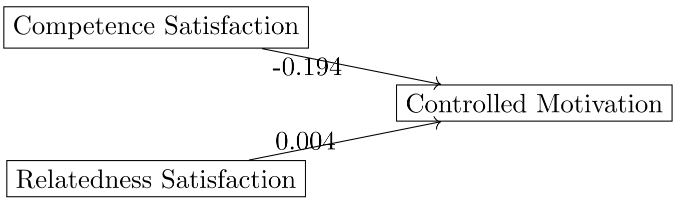

Figure 17: Graphical Presentation of the model for Hypothesis 4

We conclude that even when we consider Competence and Relatedness Satisfaction and Controlled Motivation separately, we do not obtain significant results.
We must note, though, that the RMSEA is again too high, making it more difficult to interpret.

Hypothesis 5: Competence and Relatedness Frustration are negatively related to Autonomous Motivation, but less strongly than the relationships between Competence and Relatedness Satisfaction and Autonomous Motivation, and Competence and Relatedness Frustration and Controlled Motivation

Based on the results corresponding to Figure 12 we obtain the following results:
We start by looking at the relationship between Competence and Relatedness Frustration and Autonomous Motivation and find that the standardized regression coefficients are equal to $\beta = -0.34, p > 0.05$ for Relatedness Frustration and Autonomous Motivation, and $\beta = 0.22, p > 0.05$ for Competence Frustration and Autonomous Motivation. Whereas the first appears to be negative, it is not statistically significant, so there does not appear to be a relationship.
We then look at Competence and Relatedness Satisfaction and Autonomous Motivation, but find that these relationships are not statistically significant either ($\beta = 0.16, p > 0.05$ and $\beta = -0.34, p > 0.05$) respectively.
Finally, the relationships with Controlled Motivation are also not significant: $\beta = 0.22, p > 0.05$ and $\beta = -0.05, p > 0.05$ for Competence and Relatedness Frustration respectively.

We conclude that there is no support for this hypothesis when we take into account all six Needs together.

However, when we create three separate models, shown in Figure 18, we obtain the following goodness-of-fit measures:

Table 26: Goodness-of-fit measures for Hypothesis 5

Goodness-of-fit measure	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
1: CF, RF, AM	0	0.966	0.959	(0.062,0.091)	0.073
2: CS, RS, AM	0	0.972	0.966	(0.064,0.093)	0.066
3: CF, RF, CM	0	0.960	0.952	(0.058,0.088)	0.080

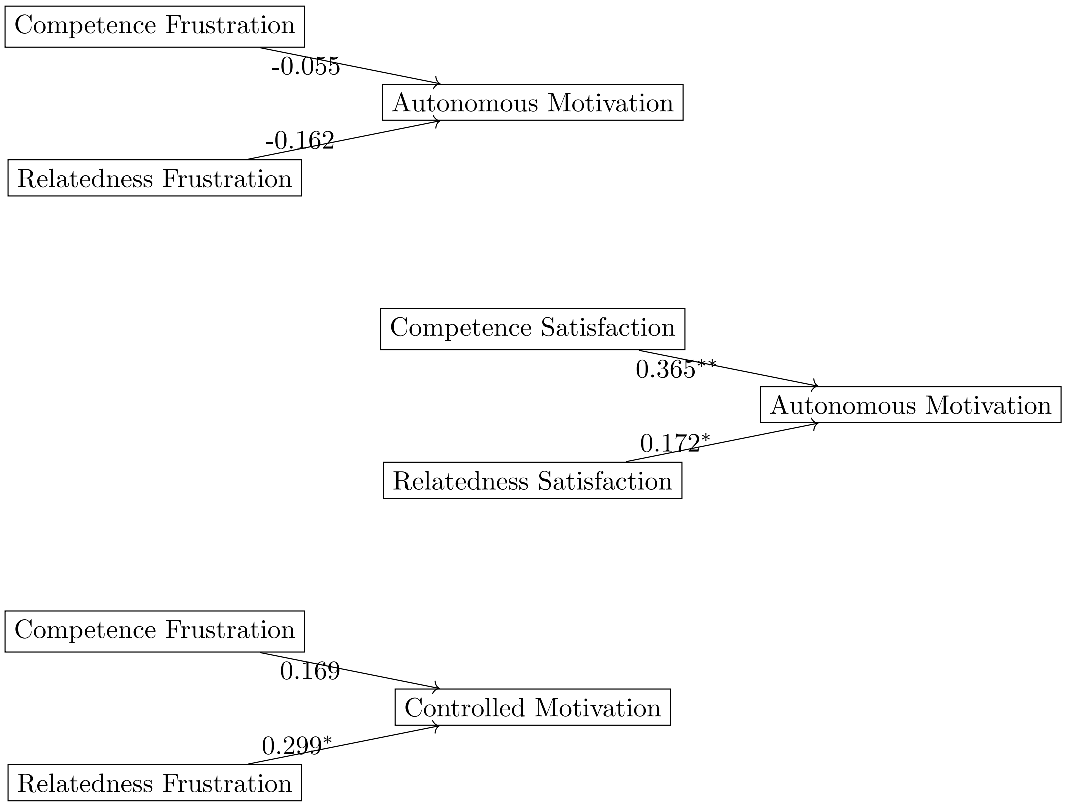

Figure 18: Graphical Presentation of the three models for Hypothesis 5

Note that the $t-$rule again holds, as the equations for all three are the same as the one for hypothesis 4 (equation 25).
Again, the results shown in Figure 18 need to be interpreted carefully, as the RMSEA values are higher than 0.06 with very high probability.

However, the results appear to be more in line with the expectations than we thought based on the full model.
We obtain positive relationships between Competence Satisfaction and Autonomous Motivation ($\beta = 0.365, p \approx 0.000$), between Relatedness Satisfaction and Autonomous Motivation ($\beta = 0.172, p \approx 0.044$, and finally between Relatedness Frustration and Controlled Motivation ($\beta = 0.299, p \approx 0.009$).
The other relationships appear to be non-significant.

Hypothesis 6: Competence and Relatedness Satisfaction positively mediate the relationship of perceived teacher cues and Autonomous Motivation

This hypothesis cannot be tested based on the data we have so far, as we have not considered the relationship between Monitoring and Scaffolding and (Autonomous) Motivation.

However, simple models can be created to test this hypothesis.
Note that in this hypothesis, a mediation model is assumed to explain the data. Recall that an explanation on mediation was given in Chapter 2.2.2.

Since we are interested in two independent variables (Monitoring and Scaffolding) and also two mediators (Competence and Relatedness Satisfaction), we will consider four separate models:

1. Model 1: Monitoring, Competence Satisfaction and Autonomous Motivation

   First, we check for identification and obtain

$$\frac{1}{2}(24)(25) = 300 \geq 51 = 2(24) + 3 \tag{26}$$

   and so the $t-$rule holds.

   The following goodness-of-fit measures were obtained for this model:

Goodness-of-fit measure	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
	0	0.981	0.979	(0.037,0.060)	0.065

Table 27: Goodness-of-fit measures for Mediation Model 1

We see that all measures indicate that the model fits the data well.

The following Figure represents the first mediation model:

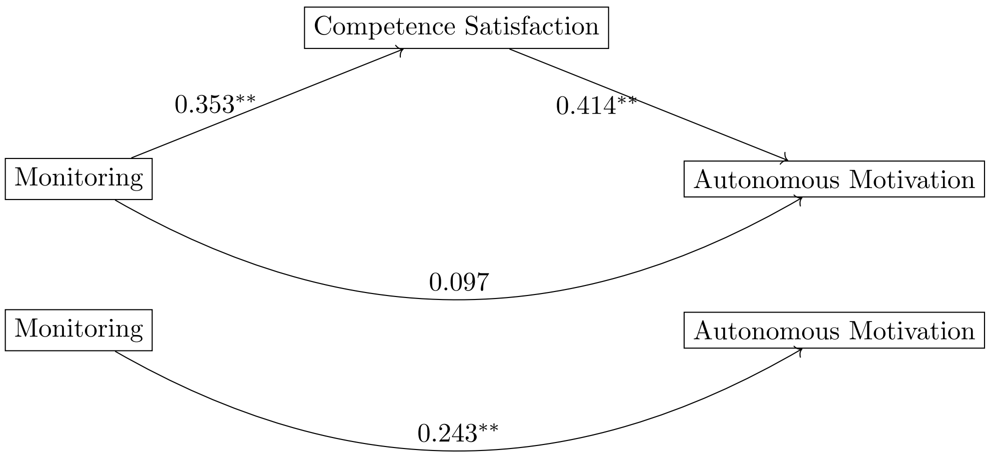

Figure 19: Graphical presentation of Mediation Model 1

We deduce from Figure 19 that we are dealing with a full mediation effect. Whereas we obtain standardized regression coefficients that are significant ($a = 0.353, p \approx 0.000$ for the relationship between Monitoring and Competence Satisfaction and $b = 0.414, p \approx 0.000$ for the relationship between Competence Satisfaction and Autonomous Motivation), the direct effect of Monitoring on Autonomous Motivation is significantly reduced to 0 (a non-significant 0.097) when including the mediator Competence Satisfaction. The mediation effect is significant: $ab = 0.146, p \approx 0.000$.

2. Model 2: Scaffolding , Competence Satisfaction and Autonomous Motivation

Again, we need to check for identification:

$$\frac{1}{2}(22)(23) = 253 \geq 47 = 2(22) + 3 \tag{27}$$

and so the $t-$rule holds.

The following goodness-of-fit measures were obtained for this model:

Goodness-of-fit measure	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
	0	0.960	0.953	(0.052,0.074)	0.076

Table 28: Goodness-of-fit measures for Mediation Model 2

We note that all measures are sufficient, except for the RMSEA, which is slightly too high. This needs to be taken into account when interpreting the following results.
The following Figure represents the second mediation model:

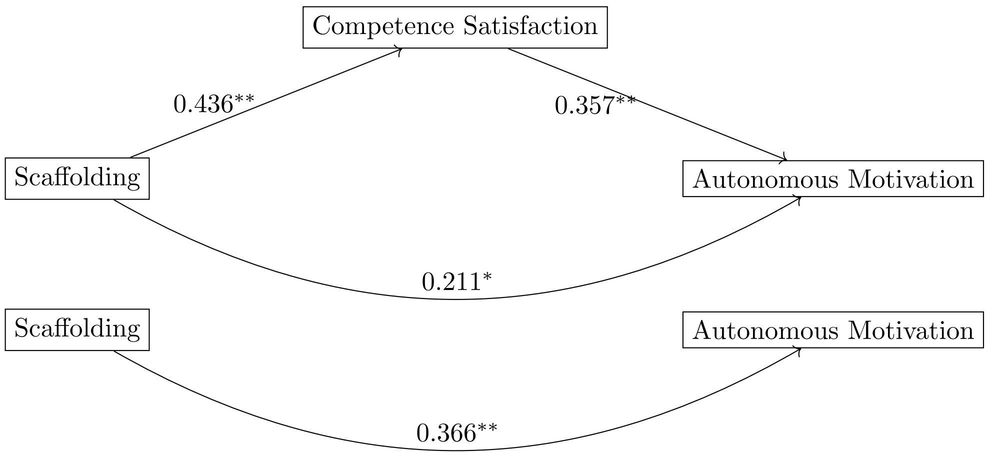

Figure 20: Graphical presentation of Mediation Model 2

We see here that we are dealing with a partial mediation effect. Whereas the direct effect of Scaffolding on Autonomous Motivation is still significant, the effect has reduced significantly from 0.366 to 0.211. We have a significant mediation effect $ab = 0.156, p \approx 0.001$.

3. Model 3: Monitoring, Relatedness Satisfaction and Autonomous Motivation

The model is similar to the first model, such that the equation for the $t-$rule is exactly the same as equation 26 and so we know that the $t-$rule holds.
The following goodness-of-fit measures were obtained for this model:

Goodness-of-fit measure	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
	0	0.983	0.981	(0.036,0.059)	0.063

Table 29: Goodness-of-fit measures for Mediation Model 1

Here we see that all measures are acceptable or good.
The following Figure represents the third mediation model:

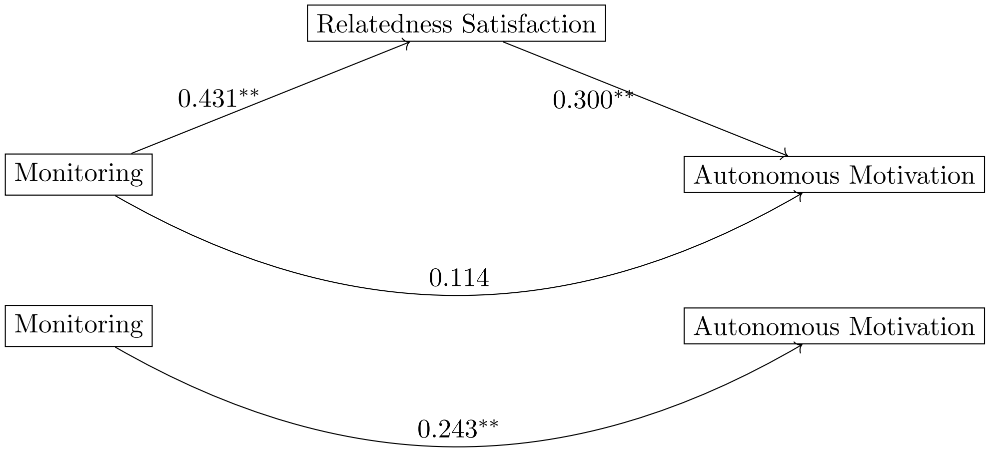

Figure 21: Graphical presentation of Mediation Model 3

Similar to what we obtained in the first mediation model, we again have a full mediation effect. The direct effect of Monitoring on Autonomous Motivation has reduced to 0 (0.114, $p > 0.05$) by including the mediator Relatedness Satisfaction. Our mediation effect is significant ($ab = 0.129, p \approx 0.002$).

4. Model 4: Scaffolding , Relatedness Satisfaction and Autonomous Motivation

   The model is similar to the second model, such that the equation for the $t-$rule is exactly the same as equation 27 and so we know that the $t-$rule holds.
   The following goodness-of-fit measures were obtained for this model:

Goodness-of-fit measure	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
	0	0.966	0.960	(0.050,0.073)	0.075

Table 30: Goodness-of-fit measures for Mediation Model 4

Unfortunately, the RMSEA is not sufficient. This needs to be considered when interpreting the following results.
The following Figure represents the fourth mediation model:

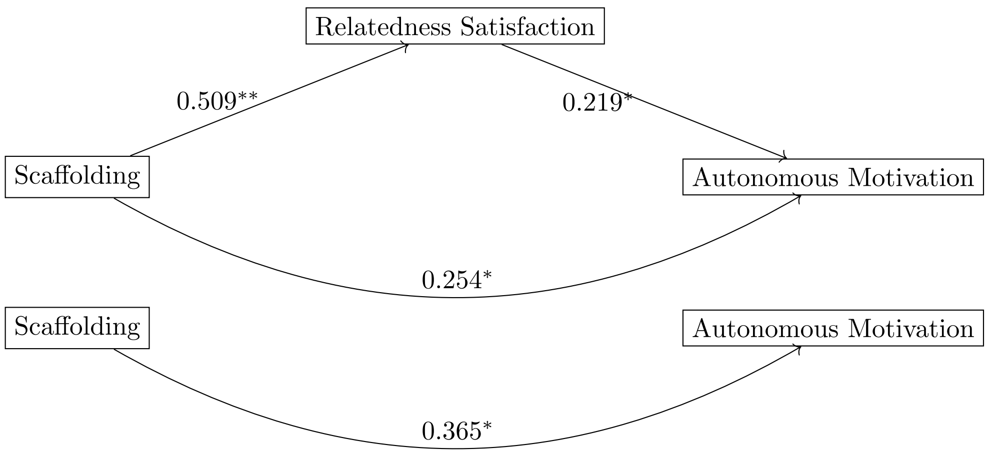

Figure 22: Graphical presentation of Mediation Model 4

Similar to what we obtained in the second mediation model, we are dealing with a partial mediation. The direct effect from Scaffolding on Autonomous Motivation has reduced significantly from 0.365 to 0.254. In spite of the significant mediation effect $ab = 0.111, p \approx 0.028$, the direct effect is, however, still significant.

To summarize, we can conclude that we can confirm this hypothesis based on our results. For Monitoring, we obtained full mediation effects when taking into account both mediators in mediation models 1 and 3. For Scaffolding , we obtained partial mediation effects in models 2 and 4. In these two models, the RMSEA was unfortunately not sufficient, making these results slightly less reliable.

Hypothesis 7: Competence and Relatedness Frustration negatively mediate the relationship of perceived teacher cues and Controlled Motivation

Again, this hypothesis cannot be tested based on the data we have so far, as we have not considered the relationship between Monitoring and Scaffolding and (Controlled) Motivation.
Once again, simple models can be created to test this hypothesis, similar to what was done for the previous hypothesis.

Since we are interested in two independent variables (Monitoring and Scaffolding) and also two mediators (Competence and Relatedness Frustration), we will consider four separate models:

1. Model 1: Monitoring, Competence Frustration and Controlled Motivation

   The model is similar to the first model from hypothesis 6, such that the equation for the $t-$rule is exactly the same as equation 26 and so we know that the $t-$rule holds.
   The following goodness-of-fit measures were obtained for this model:

Goodness-of-fit measure	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
	0	0.983	0.981	(0.032,0.056)	0.071

Table 31: Goodness-of-fit measures for Mediation Model 1

We see that all measures indicate that the model fits the data well.

The following Figure represents the first mediation model:

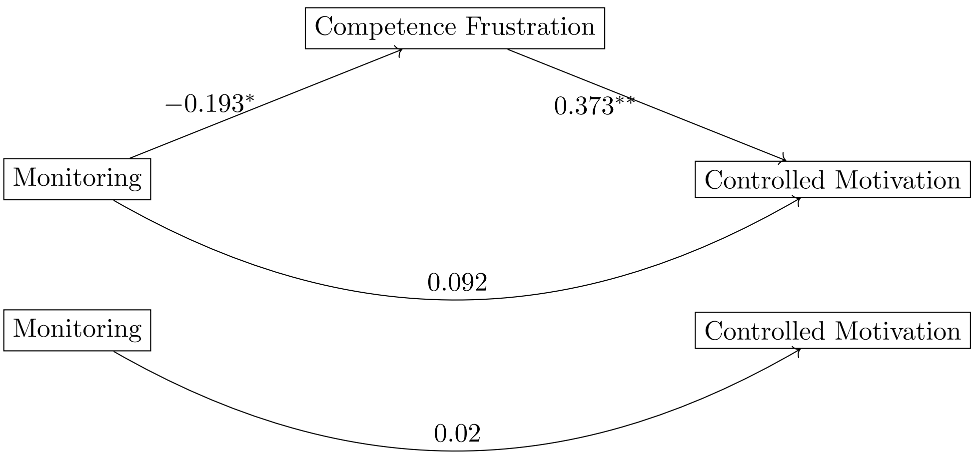

Figure 23: Graphical presentation of Mediation Model 1

First of all, we deduce from Figure 23 that there appears to be a negative mediation effect, as the total effect is in fact smaller than the direct effect. Note, however, that both the total and the direct effect

are not significant. But what does this mean? It appears that we are dealing with a special case of a mediator: our mediator is in fact a suppressor. This can happen when a and b have different signs (in this case, $a < 0$ and $b > 0$) and the direct effect is positive.
The total effect of Monitoring on Controlled Motivation is non-significant, but this is because the indirect and direct effect cancel each other out here, resulting in a small (non-significant) total effect. Even though the direct effect of Monitoring on Controlled Motivation is approximately 0 (a non-significant 0.092), it is apparently high enough to cancel out the mediation effect ($ab = -0.072, p \approx 0.043$), causing the total effect (recall that this is the sum of the indirect and direct effect) to be non-significant.

2. Model 2: Scaffolding , Competence Frustration and Controlled Motivation

   The model is similar to the second model from hypothesis 6, such that the equation for the $t-$rule is exactly the same as equation 27 and so we know that the $t-$rule holds.
   The following goodness-of-fit measures were obtained for this model:

Goodness-of-fit measure	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
	0	0.959	0.951	(0.042,0.060)	0.08

Table 32: Goodness-of-fit measures for Mediation Model 2

We note that all measures are at least acceptable.
The following Figure represents the second mediation model:

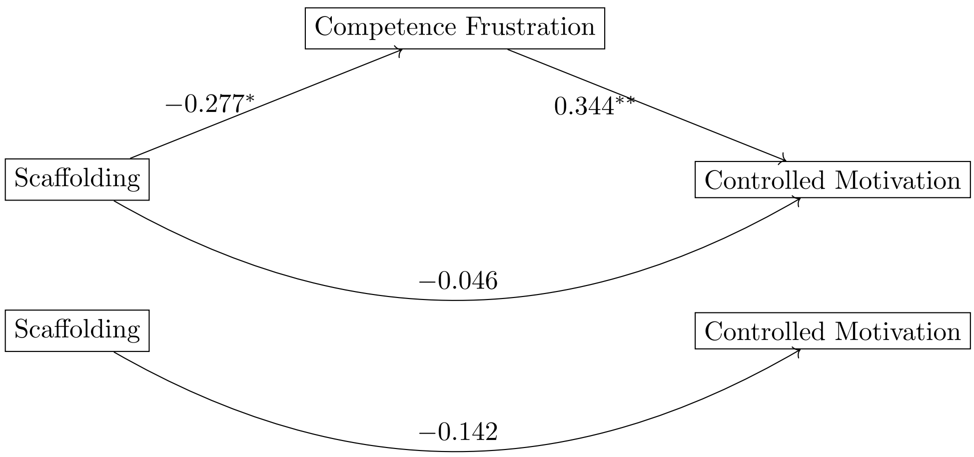

Figure 24: Graphical presentation of Mediation Model 2

We see that there is a negative relationship between Scaffolding and Competence Frustration, whereas there is a positive relationship between Competence Frustration and Controlled Motivation. Together, the indirect effect is significantly negative ($ab = -0.095, p \approx 0.025$). We see that the direct effect is also negative, though non-significant.
This results in a total effect that is (in absolute value) larger than the direct effect (note that the total effect is also not significant, though). Therefore, we cannot speak of negative mediation or suppression here.

3. Model 3: Monitoring, Relatedness Frustration and Controlled Motivation

   The model is similar to the first model from hypothesis 6, such that the equation for the $t-$rule

is exactly the same as equation 26 and so we know that the $t-$rule holds.
The following goodness-of-fit measures were obtained for this model:

Goodness-of-fit measure	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
	0	0.975	0.972	(0.042,0.063)	0.081

Table 33: Goodness-of-fit measures for Mediation Model 1

Unfortunately, not all measures are sufficient. The RMSEA is slightly too high, and the same goes for the SRMR. This needs to be taken into account when interpreting the following results.

The following Figure represents the third mediation model:

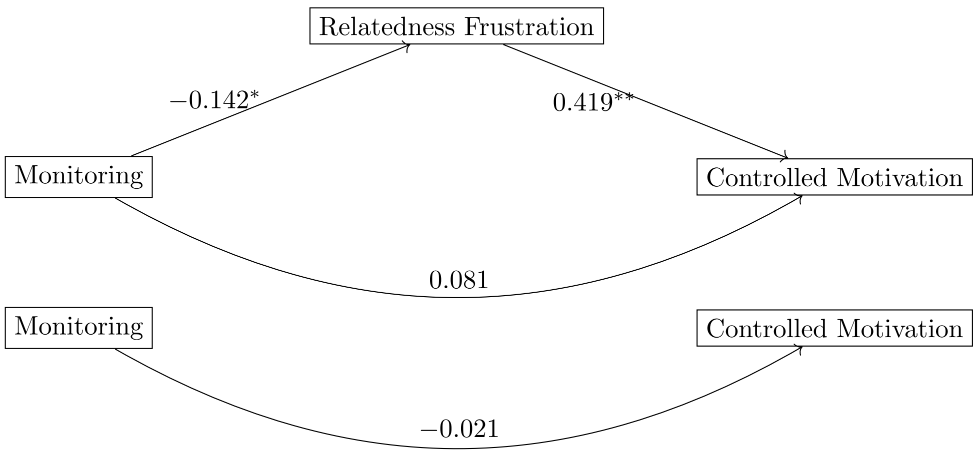

Figure 25: Graphical presentation of Mediation Model 3

We deduce from Figure 25 that we have a negative relationship between Monitoring and Relatedness Frustration, and a positive relationship between Relatedness Frustration and Controlled Motivation. Also, the indirect or mediation effect is significant: $ab = -0.06, p \approx 0.034$.
This result, combined with our finding that the total effect is smaller than the direct effect, indicates that we can indeed speak of a negative mediation effect, where the mediator is again a suppressor.

4. Model 4: Scaffolding , Relatedness Frustration and Controlled Motivation

The model is similar to the second model from hypothesis 6, such that the equation for the $t-$rule is exactly the same as equation 27 and so we know that the $t-$rule holds.
The following goodness-of-fit measures were obtained for this model:

Goodness-of-fit measure	$p(\chi^2)$	CFI	TLI	RMSEA	SRMR
	0	0.963	0.956	(0.035,0.060)	0.078

Table 34: Goodness-of-fit measures for Mediation Model 4

We see that all goodness-of-fit measures are sufficient.

The following Figure represents the fourth mediation model:

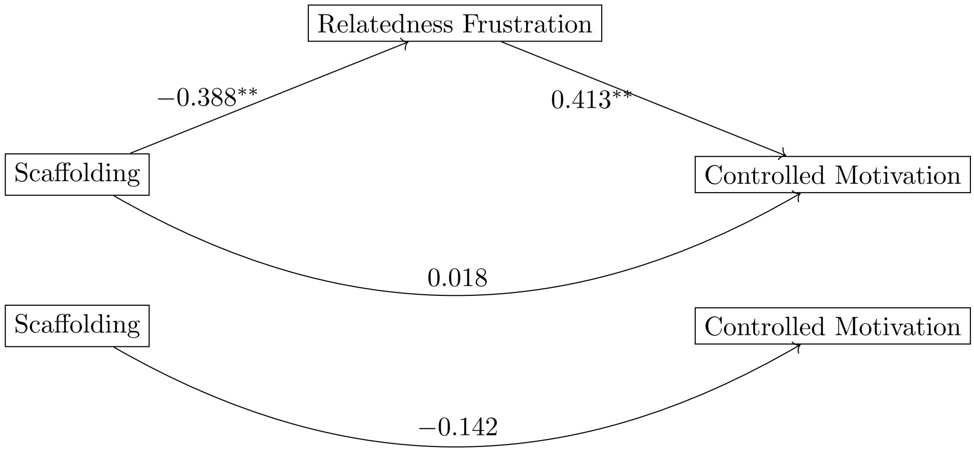

Figure 26: Graphical presentation of Mediation Model 4

We see that this model is similar to the second mediation model where considered Scaffolding and Competence Frustration. We obtain a negative indirect effect $ab = -0.16, p \approx 0.008$, but the direct effect is almost equal to 0. Also, the total effect is larger (in absolute value) than the direct effect, even if it is non-significant. Therefore, we cannot speak of negative mediation here.

To summarize, we can conclude that we can in part confirm this hypothesis based on our results. For Monitoring, we obtained negative (suppressor) mediation effects when taking into account both mediators in mediation models 1 and 3. For Scaffolding , we obtained no negative mediation effects in models 2 and 4. In these two models, the effect appears to be closer to positive mediation (as the direct effects appear to be smaller than the total effects in absolute value), but this is not supported by the significance tests.

Hypothesis 8: There is a positive relationship between Autonomous Motivation and performance and a negative, but less strong, relationship between Controlled Motivation and performance

This hypothesis can be answered directly from the model shown in Figure 13. We obtain a positive and significant standardized regression coefficient for Autonomous Motivation and Academic Performance ($\beta = 0.79, p \approx 0.001$, but a non-significant coefficient for Controlled Motivation and Academic Performance ($\beta = 0.29, p > 0.05$).
Therefore, we conclude that Controlled Motivation indeed has a less strong effect on Academic Performance compared to Autonomous Motivation. Though non-significant, it is definitely not negative, suggesting that Controlled Motivation might have a positive influence on Academic Performance, though this is not supported by our data at this point.

8 Conclusion

8.1 Summary of our results

To summarize our results, we will make a distinction between the 'journey' of creating the models, and the actual content of the models.

8.1.1 Creating the models

When it comes to creating the model, which was the main focus of this research, we can conclude that it was rather challenging to obtain a decent model for this specific subject. Fitting a mathematically correct model to non-directly observable data is not very straightforward. Mathematicians are used to working with explicit, directly observable data and are therefore not used to working with more complicated constructs that are not directly measurable. On the other hand, social scientists are used to working with implicit, non-directly observable constructs, but are less familiar with the mathematics to fit a model. This research was designed to find a solution for this problem, by creating a mathematically decent model for latent variables.

However, the challenges we faced during the process of creating a decent model for our data, were enormous.

First of all, we needed to meet the assumptions for applying SEM (Structural Equation Modelling), which was not the case. The data was ordinal and therefore automatically not normally distributed as it should have been. Also, there was a lot of missing data. These issues needed to be dealt with beforehand.
We discovered that it was possible to adjust the estimation method in `Lavaan` in `R` from Maximum Likelihood to Diagonally Weighted Least Squares, after which it became possible to work with ordinal data.
The problem of the missing values was solved too. For example, some students appeared to not have answered seriously and some specific items were not clear for many students. These were removed from the data. After this, analyses were done to make sure the remaining missing data was missing at random (MAR). This turned out to be the case, making it possible to work with this data.

When we created the model in `R` based on the relationships hypothesized, we initially obtained warnings that standard errors could not be computed and the model may not be specified. Up to this point, it is still unclear why this was the case and it will be a point of discussion in Chapter 9. Also, the goodness-of-fit measures were insufficient, indicating a bad fit.
Because it was difficult to find the cause for these problems based on such a large model, we decided to split the model into three smaller sub-models.

One of these sub-models involved the teacher cues and the six Needs.
Here, we discovered that we needed to make some improvements to the model by adding some regressions (allowing certain items to be strongly correlated to each other). We used the MI (Modification Index) to decide what items were correlated the most and manually checked whether it made sense to add a regression between these items (were the items very similar, or did it seem logical that students would answer them similarly?). After adding some regressions, most indices had improved, but the SRMR remained far too high. Therefore, we focused on individual items and removed the ones with the highest mean residual covariance compared to all other items. Six items were removed after which all goodness-of-fit measures were within the desired ranges.

For the second sub-model, involving the Needs and Motivation, only two regressions needed to be added, and no items needed to be removed. This part of the model was initially the best from the start and needed the smallest number of adjustments in order to obtain a good fit.

The third sub-model was about Motivation and Academic Performance. Here, we got the same error again as we started with: the model may not be specified as standard errors could not be

computed. We decided to use a different method to obtain standard errors: the Bootstrap method. This way, it was possible to obtain standard errors, without which it was not possible to draw any conclusions from the output.
The result was that we needed to add some regressions again, but still did not obtain the desired goodness-of-fit. Therefore, we decided to adjust the scale of Academic Performance. Initially, this one was not ordinal, like the other constructs, but simply the final grade for the course. We tried several ways to define Academic Performance and concluded that the best results were obtained when we used four quantiles and set the missing values to 0.

After we succeeded in obtaining goodness-of-fit measures as desired for all three sub-models, we decided to combine them (including all modifications) to create a final model. The results were much better than initially, but the RMSEA and the SRMR were still insufficient though. However, we did decide to share the results for this model, as it was interesting to compare to the sub-models. Since the goodness-of-fit wasn't optimal, it is important to keep this in mind when interpreting the output.

When we aimed to test the formulated hypotheses, we discovered that it was necessary to create even more (sub-)models, because the existing models were not appropriate to use as they, for example, involve more variables than the hypothesis was about or they included mediation effects we were not interested in, making direct relationships non-observable.
After creating the appropriate sub-models for the hypotheses, we discovered that it was often the case that at least one of the goodness-of-fit measures was not sufficient (most often the RMSEA).
As it is at this time not clear how exactly this affects our interpretation of the results, we thought it was nevertheless interesting to say something about them.

To conclude, it was a challenging journey to obtain decent models for our data, as often it was the case that at least one of the measures indicated a sub-optimal goodness-of-fit.
Adding regressions and removing items with high residuals seemed to be a decent solution sometimes, but for more complicated models (such as the total model, containing all latent variables) we did not succeed.
It appears to be the case that our Structural Equation Modelling worked best on the three sub-models. Perhaps the total model involved too many complicated data structures (with many mediation effects), resulting in a sub-optimal fit. Also, the small models that were created to test the hypotheses, were often slightly insufficient with respect to the RMSEA, which may be because this time there were not enough variables. These issues could be items for future research.

8.1.2 The content of the models

Another important part of this research revolves around the specific subject of our data from the survey, and trying to find relationships between the teacher cues 'Monitoring' and 'Scaffolding', the six Needs 'Competence Satisfaction', 'Competence Frustration', 'Relatedness Satisfaction', 'Relatedness Frustration', 'Autonomy Satisfaction' and 'Autonomy Frustration', the two types of Motivation 'Autonomous Motivation' and 'Controlled Motivation', and finally 'Academic Performance.'

First of all, a choice was made to make a distinction between Monitoring and Scaffolding. Initially, they were grouped together as 'Teacher Cues', but since we noticed that students' answers differed between the two constructs and it was, from the beginning, clear which items were related to Monitoring and which to Scaffolding, it seemed wise to separate the two.
Also, in previous research, factor analysis indicated that two factors were present in the thirty initial items involving Monitoring and Scaffolding.

The first sub-model, involving the perceived Teacher Cues and the four Needs: Satisfaction and Frustration for both Relatedness and Competence, shows us that Monitoring has a positive influence on Relatedness Satisfaction, and no significant influence on the other three. Scaffolding, however, has a positive influence on both Relatedness and Competence Satisfaction, and a negative influence on both Relatedness and Competence Frustration.

The second sub-model, involving all six Needs and both types of Motivation, shows us that there is a strong relationship between Autonomy Satisfaction and Autonomous Motivation, but all other relationships are non-significant.
Perhaps this means that when we know something about Autonomy Satisfaction, this is enough to make a prediction about Autonomous Motivation, without considering the other five Needs. However, it is still possible that the other five Needs are significantly related to Motivation when we do not include Autonomy Satisfaction in the model. Indeed, this turned out to be the case when we considered only a small number of Needs or only one type of Motivation for the purpose of testing some of the hypotheses. For example, Competence and Relatedness Satisfaction both have a positive influence on Autonomous Motivation, when we exclude Autonomy Satisfaction.
Also, the result that Autonomy Satisfaction has a strong positive influence on Autonomous Motivation was confirmed again by the results of the smaller model that was created to test the second hypothesis. In total, we can conclude that Autonomous Motivation is positively influenced by all three types of satisfaction. Interestingly, out of the three types of frustration, only Autonomy Frustration appeared to have a negative influence on Autonomous Motivation.
When it comes to Controlled Motivation, it appears to be the case that Autonomy and Relatedness Frustration have a positive influence, but Competence Frustration does not have any effect on it (however, as we will discuss later, when testing one of the mediation hypotheses, Competence Frustration did have a positive influence on Controlled Motivation when adjusted for Monitoring).

For the three types of satisfaction, we conclude that none of them have any influence on Controlled Motivation, even though negative influences were expected.

The third sub-model was about Motivation and Academic Performance, and shows us that Autonomous Motivation has a strong positive relationship with Academic Performance, and Controlled Motivation does not. However, it is again possible that Controlled Motivation has an influence on Academic Performance when we do not include Autonomous Motivation.

When we aimed to test the hypotheses, we came across two hypotheses that involved mediation effects. In total, eight mediation models were created, and the following interesting results were obtained:

- Competence Satisfaction fully mediates the relationship between Monitoring and Autonomous Motivation: Monitoring positively influences Autonomous Motivation, but when controlling for Competence Satisfaction there is no direct effect.

- Competence Satisfaction partially mediates the relationship between Scaffolding and Autonomous Motivation: the positive influence from Scaffolding to Autonomous Motivation is reduced when controlling for Competence Satisfaction.

- Relatedness Satisfaction fully mediates the relationship between Monitoring and Autonomous Motivation: Monitoring positively influences Autonomous Motivation, but when controlling for Relatedness Satisfaction there is no direct effect.

- Relatedness Satisfaction partially mediates the relationship between Scaffolding and Autonomous Motivation: the positive influence from Scaffolding to Autonomous Motivation is reduced when controlling for Relatedness Satisfaction.

- Competence Frustration suppresses the relationship between Monitoring and Controlled Motivation: there appears to be no relationship between Monitoring and Controlled Motivation, but this seems to be caused by the fact that the indirect and the direct effect cancel each other out.

- It is unclear how to interpret whether Competence Frustration mediates the relationship between Scaffolding and Controlled Motivation, as both the direct and the total effect are non-significant, but the indirect effect or mediation effect is significant.
- Relatedness Frustration suppresses the relationship between Monitoring and Controlled Motivation: there appears to be no relationship between Monitoring and Controlled Motivation, but this seems to be caused by the fact that the indirect and the direct effect cancel each other out.
- It is unclear how to interpret whether Relatedness Frustration mediates the relationship between Scaffolding and Controlled Motivation, as both the direct and the total effect are non-significant, but the indirect effect or mediation effect is significant.

8.2 A comparison to previous research with the same data set

In this subsection, a comparison is made between the results of this research and the ones from previous research done by two other TU-Delft students who used a similar data set (the same data set initially, but probably with some differences as we may have cleaned the data differently).

1. One of these researches was also about estimating links between latent variables using SEM by Plomp [20]. Here, the author faced similar issues with obtaining a decent fit. She also used the Modification Index to improve the model, with a final result of a sub-optimal goodness-of-fit though. Comparison possibilities are limited, as the author decided not to split Monitoring and Scaffolding like we did. Also, she did not use the Diagonally Weighted Least Squares estimator, but the Maximum Likelihood estimator to obtain standardized regression coefficients. Finally, her coefficients were not tested for significance, and therefore we cannot be sure whether the relationships found are meaningful. However, we will still report the most clear similarities and differences between the two researches:

    - When it comes to the relationships between the teacher cues and the six Needs, a positive relationship between both teacher cues and Relatedness Satisfaction was found in both researches. Also, a negative relationship was found between Scaffolding (in [20], teacher cues) and Competence Frustration. Other than this, the results differed.
    - For the six Needs and Motivation, both researches obtained (strong) relationships between Autonomy, Relatedness and Competence Satisfaction and Autonomous Motivation. Also, Autonomy Frustration has a positive influence on Controlled Motivation and a negative influence on Autonomous Motivation in both researches.
    - Finally, when it comes to the relationships between Motivation and Academic Performance, both researches found a positive relationship between Autonomous Motivation and Academic Performance.
      Plomp also found a negative relationship between Controlled Motivation and Academic Performance, but we found they were not significantly related.

    To conclude, it is difficult to compare the results because the analyses were done very differently. However, it is interesting to see there are still some similarities.

2. The other research we make a comparison with, was done by Brouwer [4]. She also used the same data set, and also applied SEM on it, with the largest difference that she used a Bayesian variant of SEM. Unfortunately, Brouwer also faced issues obtaining a decent goodness-of-fit for her model, and worked with sub-optimal results. Here, a comparison is again not straightforward, as Brouwer did not separate Monitoring and Scaffolding like we did, and it is not clear whether the fact that we are dealing with ordinal data was taken into account.
   Finally, the reported results are based on the standardized coefficient estimates but no report was made about the statistical significance of the results. Nonetheless, it is interesting to report some similarities and differences:

   - For the relationships between the Teacher Cues and the Needs, the relationships obtained are similar, with the exception that Brouwer obtained a positive relationship between teacher cues and Competence Frustration, where we obtained negative ones (with statistical significance only for Scaffolding).

   - When it comes to the relationships between the Needs and Motivation, we obtained that all three types of Satisfaction are positively related to Autonomous Motivation, but in [4] only Competence Satisfaction had a positive relationship with Autonomous Motivation. For Relatedness Satisfaction, there was no relationship and for Autonomy Satisfaction, the coefficient was negative even. This is a remarkable difference, as we obtained a very strong positive relationship between Autonomy Satisfaction and Autonomous Motivation. For Controlled Motivation, the relationships obtained also differ greatly.

   - Finally, a positive relationship between Autonomous Motivation and Academic Performance was obtained in both researches. Brouwer also found a positive relationship between Controlled Motivation and Academic Performance (we did too, but the relationship was not statistically significant).

   To conclude, there are some similarities but also large differences between the two researches.

In total, we conclude that comparisons between all three researches are difficult to make. What stands out is that all three researchers obtained a strong positive relationship between Autonomous Motivation and Academic Performance. All other relationships are less obvious as large differences were obtained between all researches.

9 Discussion

To summarize, it took some time and effort to obtain a decent model and therefore obtain results that could be interpreted. The 'problems' of the ordinal structure of the data and the fact that there were many missing values needed to be handled first. The model needed to be split up into smaller sub-models in order to analyze the relationships between latent variables better. Adjustments to the models were made by adding regressions and deleting some items. For the complete model, bootstrapping was necessary to obtain standard errors.

With respect to the content of the models, some interesting results were obtained. The most interesting result was that when we consider Academic Performance, it appears to be that Autonomous Motivation is of positive influence. Autonomous Motivation, in turn, was positively influenced by Autonomy, Competence and Relatedness Satisfaction, with the first having the strongest influence. Autonomy Frustration appears to be the only type of Frustration having a negative effect on Autonomous Motivation. Both Monitoring and Scaffolding appear to have a positive influence on Relatedness Satisfaction, but Scaffolding also has a positive effect on Competence Satisfaction, and a negative effect on Relatedness and Competence Frustration.
Both Competence and Relatedness Satisfaction fully mediate the relationship between Monitoring and Autonomous Motivation, and they both partially mediate the relationship between Scaffolding and Autonomous Motivation.

There are, however, some points to take into consideration when interpreting the results:
Before we analyzed our data, some choices had to be made. One choice that was made was the splitting of 'Teacher Cues' into 'Monitoring' and 'Scaffolding.' Reasons to do this were that we noticed that students answered differed between the two constructs (in previous research, factor analysis indicated that there were indeed two factors present in the thirty initial items involving Monitoring and Scaffolding). However, it could be interesting to see how the models would change if we considered 'Teacher Cues' as one latent variable. The reason this was not investigated in this research was because there were already a lot of models to consider and doing this would almost double the number of total models analyzed.
Other choices that we expect could be of great influence are the choices that were made when cleaning the data. These include the removal of some students, but perhaps more importantly adjusting some of the data as students answered '2.5' instead of either 2 or 3, for example. As was explained in Chapter 5, we decided to round down answers below 3 and round up answers above 3 to make the answers less neutral. Other options would have been to round every answer either up or down, or simply remove these students entirely. It is not clear how much influence this choice has on the results, but as some latent variables are made up of only four observed variables, it is something to take into account.

In this research, there are also some shortcomings worth mentioning. One shortcoming that stands out most, perhaps, is that when we created the complete model in `R` based on the relationships hypothesized, we initially obtained warnings that standard errors could not be computed and the model may not be specified. Also, the goodness-of-fit measures were insufficient, indicating a bad fit.
It is unclear why these standard errors could not be computed. It is clear, however, that it is very hard to interpret the results of such a large model to begin with. In the complete model, there is a large number of mediation effects, resulting in very different results than when sub-models are considered or when relationships between only a small number of latent variables are analyzed, as we did to test for mediation effects, for example. So even though we found a way around this shortcoming by using Bootstrapping to compute standard errors, one can wonder about the usefulness of this complete model to begin with. Also, considering the fact that Bootstrapping was used only for the complete model and not for every model (Bootstrapping on this large data set turned out to be a very time-intensive procedure), it might be difficult to compare this model to the others as well.
We had to make a decision between making use of Bootstrapping and finding some interpretable results this way, or not making use of Bootstrapping and concluding that the complete model simply did not represent the data well. We decided to make use of Bootstrapping, but because of this shortcoming, we also decided to create sub-models, which turned out to be a good idea in terms of being able to interpret the results. It appeared to be the case that our Structural Equation Modelling worked best on the

three sub-models and the Mediation models. Perhaps the total model involved too many complicated data structures (with many mediation effects), resulting in a sub-optimal fit, even if it was possible to compute standard errors.

Another shortcoming of this research is that one item was removed from the variable 'Monitoring', based on too many missing values. Previous research has shown that in the original data set, the consistency measured by Cronbach's Alpha was very good, but this was not tested in the current research where one item was removed. We don't expect this of be much influence, though, as it was only one item out of 16 items total.

Something else that is worth mentioning is that the evaluation with respect to the usefulness of a model, based on the goodness-of-fit measures, depends heavily on the interpretation of the user. Table 1 in Section 2.5 provides a commonly used interpretation for the goodness-of-fit measures, but not all researchers agree on the cutoff-values presented here. More importantly, what does it mean if one of the goodness-of-fit measures is slightly off, like the RMSEA for example? It doesn't immediately make the model useless.
In this research, models were adjusted in order to obtain the best possible goodness-of-fit measures, and when one or more gof measures were still slightly out of bounds, this was mentioned, as it is something to consider when interpreting the results.

Other influences on the goodness-of-fit measures are the number of used variables, the sample size, and the type of estimator that was used in order to obtain the results. Shi, Lee and Maydeu-Olivares discovered, for example, that when a higher number of variables were used, this affected the CFI, TLI and the RMSEA in a negative way, suggesting a worse fit [24]. They also found this was the case for a smaller sample size.
Xia and Yang claim that the conventional cut-off values as proposed in table 1 in Section 2.5, should only be applied when the conventional estimator (the Maximum Likelihood Estimator) is used. They made a comparison between the ML and the DWLS estimators and claim that for the DWLS estimator, the CFI and TLI always tend to be higher and the RMSEA always tends to be lower (in all cases suggesting a better fit) [30]. They argue that using the same cutoff-values when the Diagonally Weighted Least Squares Estimator is used, is therefore not recommended, as there is a tendency not to discover model-data misfit. However, up to today there is no alternative to the cutoff-values from table 1, and therefore we decided to use them in this research anyway. But it has become clear that finding a good model fit is not that easy or straightforward.

Disagreement among researchers also exists when it comes to making use of the Modification Index in order to improve the model-fit. MacCallum, Roznowski and Necowitz argue that modification of the model based on the MI will always result in better model-fit, but this doesn't necessarily mean the model is actually better [15]. They claim this way of adjusting the model is too data-driven not based on theory, and this type of adjustment has a tendency to make the model fit better based on this specific sample and its 'chance characteristics,' which therefore not necessarily makes the model better for the population.
In this research, we decided to make use of the Modification Index, but only for regressions that made sense to us theoretically as well. All suggested modifications were manually checked and were only added when it was understandable why students would answer them in a similar fashion.

Other than adressing some shortcomings of this research, there are also some new ideas for future research. One could be to make use of the demographic data more when analyzing the results. We know, for example, the age and gender of most of the students, but this information was not used in this research. It could be interesting to find out if there is a gender difference when considering the relationships between our latent variables. What if the relationship between Autonomous Motivation and Academic Performance is stronger for females than for males, for example? This could be interesting information. In this research, however, only 24 females participated, whereas there were 152 males. In future research, if more students participate in the survey, it could be interesting to investigate further.

To conclude, we aim to interpret the results for education purposes: what connections can we make between the results of this research and the way education leads to better Academic Performance of students? The most obvious conclusion is that Autonomous Motivation plays a big, positive role in Academic Performance. Therefore, if we want better Academic Performance in our students, we need to focus on improving Autonomous Motivation. It turns out that intrinsic motivation (wanting to learn, enjoying studying in general, etc.) is much more important than external motivation (feeling guilty about not studying, living up to other people's expectations, etc.) The focus should therefore be on Autonomous Motivation.
Our results show us that Autonomous Motivation is influenced by Autonomy Satisfaction most, and in a positive way: if students experience a sense of choice and freedom in their undertakings, feel that their decisions about course activities reflect what they really want and express who they really are, and are really interested in what they are doing, this has a very positive influence on Autonomous Motivation. Similarly, Autonomy Frustration appears to have a very negative influence on Autonomous Motivation. These feelings of Autonomy might be difficult to influence for a teacher, however. If we leave out Autonomy Satisfaction, we find that Relatedness Satisfaction and Competence Satisfaction also play a positive role: feeling connected to other students, and feeling competent to achieving their goals can also be of importance in being intrinsically motivated, albeit to a lesser extent. However, these two constructs can be influenced by teachers, and are therefore valuable to look into more: How can the way education is given, influence these types of Satisfaction? Our results show us that Monitoring can influence Relatedness Satisfaction and Competence Satisfaction in a positive way: it helps if the lecturer encourages students to reflect, stimulates to think about improvement, discusses progress that was made, etc. Monitoring has no effect on Autonomous Motivation directly, as it turns out that Relatedness and Competence Satisfaction both fully mediate this relationship.
We find that Scaffolding plays an even bigger role than Monitoring when it comes to Relatedness and Competence Satisfaction: it helps if the lecturer is flexible in explaining topics in several ways, if there are opportunities to share and discuss what is learned, if contribution to the lessons and interaction between classmates is allowed, if the student feels that the assignments reflect the material well, etc.
Scaffolding even turns out to affect Autonomous Motivation in a direct way, as becomes clear from the partial mediation effects: Relatedness and Competence Satisfaction explain the positive relationship between Scaffolding and Autonomous Satisfaction only partly.
We conclude that, even though Autonomy Satisfaction is very important for Autonomous Motivation, the lecturer can have the most influence on Relatedness and Competence Satisfaction, and even on Autonomous Motivation directly, by doing well on the 'Scaffolding' cues.

References

[1] Asparouhov, T. & Muthén, B. (2010). Simple second order chi-square correction. Retrieved from www.statmodel.com/download/ WLSMV$_{n}ew_{c}hi21.pdf$

[2] Barrett, P. (2007). Structural equation modelling: adjudging model fit. *Personality and Individual Differences, 42,* 815–824.

[3] Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588-600.

[4] Brouwer, A.C. (2021). Bayesian Structural Equation Modeling explained and applied to Educational Science.

[5] Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

[6] Cochran, W. G. (1952). The Chi-square Test of Goodness of Fit. *The Annals of Mathematical Statistics. 23 (3)*: 315–345.

[7] Finney, S. J. & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. *Structural Equation Modeling: A Second Course* 269-314.

[8] Gazeloglu, C. & Greenacre, Z.A. (2020). Comparison of weighted least squares and robust estimation in structural equation modeling of ordinal categorical data with larger sample sizes. *Cumhuriyet Science Journal, 41*, 193-211.

[9] Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55.

[10] Kaplan, D. (2008). Structural Equation Modeling: Foundations and Extensions (2nd ed.)

[11] Kenny, D.A. (2020). Measuring Model Fit. [Online] Available at: http://www.davidakenny.net/cm/fit.htm [Accessed on 28 April 2021]

[12] Kline, R. B. (2010). Principles and practice of structural equation modeling (3rd Edition). New York, NY: Guilford Press.

[13] Kumar, S. & Upadhaya, G. (2017) Structure Equation Modeling Basic Assumptions and Concepts: A Novices Guide. *International Journal of Quantitative and Qualitative Research Methods, 4* 10-17.

[14] MacCallum, R. C. (2003). 2001 Presidential address: Working with imperfect models.

[15] MacCallum, R. C., Roznowski, M. & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111,* 490-504.

[16] MacKinnon, D.P. (2008), Introduction to statistical mediation analysis. Erlbaum Psych Press. *Multivariate Behavioral Research, 38,* 113-139.

[17] Marsh, H. W., Hau, K. T. & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11,* 320–341.

[18] M$\hat{i}$ndrila, D. (2010). Maximum Likelihood (ML) and Diagonally Weighted Least Squares (DWLS) Estimation Procedures: A Comparison of Estimation Bias with Ordinal and Multivariate Non-Normal Data. *International Journal of Digital Society (IJDS), 1,* 60-66.

[19] Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44,* 443–460.

[20] Plomp, V. (2020). Estimating links between latent variables using Stuctural Equation Modeling in R.

[21] PRIME, TU Delft [Online] Available at: https://www.tudelft.nl/eemcs/the-faculty/departments/applied-mathematics/education/prime/ [Accessed on 28 November 2022]

[22] Savalei, V. & Rhemtulla, M. (2013). The performance of robust test statistics with categorical data. *British Journal of Mathematical and Statistical Psychology, 66,* 201–223.

[23] Schumacker, R. E. & Lomax, R. G. (2010). A beginner's guide to structural equation modeling (3rd ed.). New York, NY: Routledge Academic.

[24] Shi, D., Lee, T. & Maydeu-Olivares, A. (2019) Understanding the Model Size Effect on SEM Fit Indices. *Educational and Psychological Measurement,79(2)* 310–334.

[25] Suhr, D. (2006) The basics of structural equation modeling. [Online] Available at: http://www.lexjansen.com/wuss/2006/tutorials/tut-suhr.pdf [Accessed on 28 April 2021].

[26] Tucker, L. R. & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1-10.

[27] Wang, J. and Wang, X. *Structural Equation Modeling: Applications Using Mplus.* Wiley, 2012

[28] West, S. G., Taylor, A. B. & Wu, W. (2012). Model fit and model selection in structural equation modeling. *In R. H. Hoyle (Ed.), Handbook of structural equation modeling* (pp. 209-231). New York, NY: Guilford Press.

[29] Whalley, B. Modification Indices, Just enough R. [Online] Available at: https://benwhalley.github.io/just-enough-r/modification-indices.html

[30] Xia, Y. & Yang, Y. (2018) RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods, 51,* 409-428.

Appendix

The appendix can be provided upon request.