Fitting statistical models for studying teachers' influence on students' math anxiety

Het passen van statistische modellen om de invloed van leraren op de wiskundeangst van leerlingen te bestuderen

T.J.M. Schelling
5980550

A thesis presented for the degree of
BSc Applied Mathematics

Statistics
Delft University of Technology
The Netherlands
June 15, 2026

Lay Summary

Math anxiety is the fear or anxiety that people experience in mathematical situations. It is an increasing problem because math-anxious people tend to avoid math in their daily life and career. Because math anxiety can discourage students from pursuing technical careers, understanding its causes is important. One of the places to have influence on math anxiety is during high school. Therefore, the influence of teachers on students' math anxiety need to be investigated. A great deal of research has already been conducted, using different approaches. The main goal of this study is to determine the best approach in researching the influence of math teachers on students' math anxiety. To find the best method, three different strategies have been tried on a small dataset. The approaches are designed to explain math anxiety through teacher characteristics, the teachers' mindset, and teaching methods. These three procedures were compared with each other. The method where math anxiety was explained per teacher and the method where math anxiety per student was investigated and explained by looking into students' interpersonal behavior with their teacher had the best ability to explain the results.

Abstract

Math anxiety, defined as fear or anxiety in math-related situations, is an increasing problem worldwide. This type of anxiety is experienced not only by students but also by adults, in situations such as paying with cash. This trend is concerning because individuals with high levels of math anxiety often avoid math-related activities, including educational and career opportunities. There is a shortage of professionals in technical fields, and this shortage will not decrease when more math anxiety is present.

Research into how to reduce the math anxiety of people begins at (high) school. Studies have confirmed that high school teachers have an influence on the increase or decrease of students' math anxiety. The key question is how this influence is expressed. Previous studies have investigated this question using a variety of statistical models. The purpose of this study is to determine which statistical model explains the most about the influence of teachers on students' math anxiety.

To investigate this question, data were collected at a high school in the Netherlands via questionnaires for teachers and students. The teacher survey contained the topics: mindset, instruction, and teacher emotions. Students completed questionnaires on their math anxiety and their perceived relationship with their teacher.

Subsequently, three models are constructed. The first model is a linear regression model where the average math anxiety of all students of one teacher is the response variable and the teacher characteristics are the explanatory variables. The second model uses the interpersonal relationship of a class with their teacher as explanatory variables and the average math anxiety per class as the response variable. This is a two-level mixed model. The last model is a three-level mixed-effects model and uses the interpersonal relationship of student and teacher as explanatory variables and the math anxiety of that student as the response variable. Three variants of the student-level model are examined: a model that has clusters on class and teacher levels. And two that only contain one cluster level, a two-level mixed model.

The models were compared using R^2 measures, the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC). The results indicate that the models based on teacher-level average anxiety and student-level anxiety with teacher clustering provided the best fit to the data. Future research should examine generalized models and validate the findings using larger datasets.

Contents

Symbols and their meaning used in the report

Symbol	Meaning
Q	The number of explanatory variables in the model
P	The number of explanatory variables in the second level of the model
R	The number of explanatory variables in the third level of the model
N	The sample size
J	The number of clusters in the second level
K	The number of clusters in the third level
n_j	The sample size of the jth cluster ($j \in \{1,...,J\}$)
n_k	The sample size of the kth cluster ($k \in \{1,...,K\}$)

Table 1: Symbols and their meaning used in the report

1
Introduction

Math anxiety has become an increasingly important educational and societal concern. Math anxiety, which is different from test anxiety, anxiety to fail, and other anxieties, is specific to mathematical situations. Math anxiety refers to feelings of fear or anxiety that arise not only during tests but also in everyday situations involving math (Luttenberger et al., 2018). Examples outside the classroom include doing math homework, paying with cash in a supermarket, or checking the bill in a restaurant.

Math anxiety often develops during childhood and can persist into adolescence and adulthood. Even primary school teachers have an influence on the development of math anxiety (Beilock and Maloney, 2015). Research has shown that high school teachers can significantly influence students' levels of math anxiety (Radišić et al., 2015). After high school, adults still experience some math anxiety. About 93% of adults in the USA experience some math anxiety, and 30% show high anxiety (Luttenberger et al., 2018).

Math anxiety is a substantial problem. It impacts more than just school results. High-anxiety students tend to avoid college majors and careers that are related to mathematics (Ashcraft, 2002). There is a growing shortage of graduates entering STEM[1]-related fields, so upcoming math anxiety will only increase this shortage dramatically (Klingberg, 2025). Furthermore, avoiding math may reinforce rather than reduce anxiety.

Understanding the causes of math anxiety is therefore essential. Most of the anxiety is already formed even before a student starts school. There is a cognitive component of math anxiety; people with high math anxiety tend to do worse in counting objects and comparing numbers and have worse spatial ability (Beilock and Maloney, 2015). There is also a role of parents in developing math anxiety. Parents who have high math anxiety are less likely to help their children with math homework (Beilock and Maloney, 2015). On the other hand, parents with high math anxiety that do help their children with homework are more likely to express their own dislike or discomfort with mathematics.

A good moment to reduce math anxiety is during high school, where teachers can influence students' anxiety. However, the mechanisms through which teachers influence students' math anxiety remain unclear. A great deal of research has been conducted to figure out what the exact role of teachers is. Studies used the influence of interpersonal student-teacher behavior (Lapointe et al., 2005), the influence of a growth/fixed mindset (Seo and Lee, 2021), the influence of teacher emotions (Frenzel et al., 2021), and the influence of concept/procedure-focused instruction (Lyu et al., 2025). These studies used heterogeneous statistical models to answer their research questions.

This pilot study was conducted in preparation for a larger study that will be conducted in November. The research will be done at the Delft University of Technology. The research in this paper contains the pilot study prior to the big research. The purpose of this pilot study is to determine which research models are best to use in the research of the influence of math teachers on students' math anxiety.

[1] Science, Technology, Engineering, Mathematics

2

Linear Mixed-Effects Models

For this research (the most useful models to use in the research study of math anxiety) several models are needed which can be compared after. In this Chapter the mathematical background of the models is discussed. The notation in this chapter is based on Fox (2016) and Snijders (2011). In Section 2.1, a comparison will be made between two different models: the linear mixed model and the linear model. This comparison is made to explain why the mixed model is used instead of the standard linear model. In Section 2.2 the characteristics, assumptions, and estimators of the two-level mixed model will be discussed. In Section 2.3 the three-level model will be discussed in detail. In the last section of this chapter (2.4), the mathematical background of the statistical tests for comparison will be explained.

2.1. Linear Mixed Models vs Linear Models

Regression models are widely used in statistical analyses. Examples include studying the influence of self-study time on students' grades, temperature on rainfall, or a giraffe's height on its weight.
In many situations, these relationships can be analyzed using a (generalized) linear model. Linear models are particularly suitable when observations can be assumed to be independent of one another. However, in many social science studies, observations are naturally grouped. For example, students may be clustered within classes, classes within school, or patients within hospitals, this is called hierarchical data. Individuals belonging to the same group often share characteristics and therefore cannot be considered fully independent.
When such clustering is present, the assumptions of a standard linear model may be violated. Mixed-effects models provide a way to account for these hierarchical data structures by incorporating both fixed effects and random effects. This section explains the differences between linear and mixed-effects models and discusses why mixed-effects models are often more appropriate for hierarchical data.

The linear model or linear regression model is given by

$$\begin{aligned} y_i &= \beta_1 + \beta_2 X_{2i} + \cdots + \beta_Q X_{Qi} + \varepsilon_i \qquad i \in \{1, \ldots, N\} \\ &= \boldsymbol{\beta}\mathbf{X}_i + \varepsilon_i \end{aligned} \tag{2.1}$$

- y_i is the response variable for a certain data point i
- β_q is the coefficient for the explanatory variable for $(q \in \{1, \ldots, Q\})$. They form the $(1 \times Q)$ vector $\boldsymbol{\beta}$
- X_{qi} is the explanatory variable for a data point i for $(q \in \{1, \ldots, Q\})$. They form the $(Q \times 1)$ vector $\mathbf{X}_i$
- ε_i is the measurement error for a data point i. We assume ε_i is normally distributed with normal distribution $\varepsilon_i \sim N(0, \sigma^2)$

In matrix-vector notation this looks like

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.2}$$

- $\mathbf{Y}$ is an $(N \times 1)$ response vector

- $\boldsymbol{\beta}$ is an $(Q \times 1)$ vector of coefficients
- $\mathbf{X}$ is an $(N \times Q)$ matrix of explanatory variables
- $\boldsymbol{\varepsilon}$ is an $(N \times 1)$ vector of errors. We assume $\boldsymbol{\varepsilon}$ is normally distributed with multivariate normal distribution $\boldsymbol{\varepsilon} \sim N_N(0, \sigma^2 \boldsymbol{I}_N)$

As mentioned earlier, we can use linear models in cases where the observations are independent. In the example case of predicting the weight of the giraffe; the response variable is the weight and the explanatory variable is either height, length or age of the giraffe.
In the researches where the observations are dependent, we cannot use the standard linear model. In the case of the giraffe, we only measure each giraffe once, which makes our observations independent. However, in most social studies, the participants are related to each other; they are in the same class, family, etc. which results in dependent data. This does not mean that the data is useless, but it has to be analyzed by another statistical model. That is why linear mixed(-effects) models (LMM) are used. The difference with the standard linear model in Equation 2.1 is an extra term for clustered data, for example students in the same class.
The equation for a single data point in the LMM is given by

$$\begin{aligned} y_i &= \beta_{1j} + \beta_{2j} X_{2ij} + \cdots + \beta_{Qj} X_{Qij} + \varepsilon_{ij} \qquad i \in \{1, \ldots, N\}, j \in \{1, \ldots, J\} \\ &= \boldsymbol{\beta}_j \mathbf{X}_{ij} + \varepsilon_{ij} \end{aligned} \tag{2.3}$$

- y_i is the response variable for a certain data point i
- β_{qj} is the coefficient for $(q \in \{1, \ldots, Q\})$. They form the $(1 \times Q)$ vector $\boldsymbol{\beta}_j$
- X_{qij} is the explanatory variable for a data point i for $(q \in \{1, \ldots, Q\})$. They form the $(Q \times 1)$ vector $\mathbf{X}_{ij}$
- ε_{ij} is the measurement error for a data point i. We assume ε_{ij} is normally distributed with normal distribution $\varepsilon_{ij} \sim N(0, \sigma_j^2)$

In the LMM equation, y_i and $\mathbf{X}_{ij}$ are equal to those in Equation 2.1. $\boldsymbol{\beta}_j$ and ε_{ij} differ from the linear model. In Section 2.2, these parameters will be explained in more detail.

2.2. Two-Level Linear Mixed Models

Linear mixed-effects models are mainly used when we think that the observations are dependent. This can be due to different reasons. Two separate cases do apply; the case when one person (or one data point) is observed multiple times, for example, when following one person during a diet; this is called longitudinal data. The other case is when multiple people (or data points) lie in the same group, for example, multiple patients who have the same doctor or therapist; this is called hierarchical data.
In the case of this research, there are multiple students in the same class who have the same teachers and courses and multiple classes that have the same teacher. So, the observed math anxiety of the students is dependent, hierarchical data. Therefore, we will only look into the details of the LMM for hierarchical data in this section.

If we look back to Equation 2.3, we see different components and different indices. In this case the i's refer to the different observations (students/classes) and the j's to the different clusters (classes/teachers). Because students in the same classroom are taught by the same teacher, and therefore the teacher influences all students, the y_i's within the same cluster are dependent. Therefore, we can assume that every cluster has their own regression coefficient, β_{qj}'s. The regression coefficient is the value for β_{qj}, which increases or decreases the math anxiety by one unit if the corresponding explanatory variable increases by one unit. These are given by

$$\beta_{qj} = \gamma_q + u_{qj} \qquad j \in \{1, \ldots, J\}, q \in \{1, \ldots, Q\} \tag{2.4}$$

- β_{qj} is the regression coefficient for the qth explanatory variable and the jth cluster
- γ_q is the fixed effect of the slope for the qth explanatory variable, independent of the cluster
- u_{qj} is the random effect of the slope for the qth explanatory variable. We assume u_{qj} is normally distributed, with normal distribution $u_{qj} \sim N(0, \tau_q)$

  - We assume that all u_{qj} and $u_{qj'}$ are independent for $j \neq j'$, because we assume that the slope is different for each cluster

Note that we do not always assume that every slope has a random component. In some cases a characteristic only has influence on one individual and not on the whole classroom. Therefore u_{qj} is not used for all regression coefficients β_{qj}. We will leave out u_{qj} in those cases, which gives $\beta_{qj} = \gamma_q$.

We can combine Equation 2.3 and Equation 2.4, to get the following equation:

$$\begin{aligned} y_i &= (\gamma_1 + u_{1j}) + (\gamma_2 + u_{2j})X_{2ij} + \cdots + (\gamma_Q + u_{Qj})X_{Qij} + \varepsilon_{ij} \qquad i \in \{1,...,N\}, j \in \{1,...,J\} \\ &= (\boldsymbol{\gamma} + \mathbf{u}_j)\mathbf{X}_{ij} + \varepsilon_{ij} \end{aligned} \tag{2.5}$$

- $\boldsymbol{\gamma}$ is an $(1 \times Q)$ vector of fixed effects
- $\mathbf{u}_j$ is an $(1 \times Q)$ vector of random effects

We will now introduce the Laird-Ware form (Laird and Ware, 1982). For the Laird-Ware form, we split Equation 2.5 into two components, the fixed component and the random component.

$$\begin{aligned} y_i &= \gamma_1 + \gamma_2 X_{2ij} + \cdots + \gamma_Q X_{Qij} + u_{1j} + u_{2j} Z_{2ij} + \cdots + u_{Pj} Z_{Pij} + \varepsilon_{ij} \qquad i \in \{1,...,N\}, j \in \{1,...,J\} \\ &= \boldsymbol{\gamma}\mathbf{X}_{ij} + \mathbf{u}_j\mathbf{Z}_{ij} + \varepsilon_{ij} \end{aligned} \tag{2.6}$$

- $\mathbf{Z}_{ij}$ is an $(P \times 1)$ vector which contains the explanatory variables only if they are modeled with a random effect
  - In some cases $\mathbf{X}_{ij} = \mathbf{Z}_{ij}$, that is when every slope has a random effect
  - In the special case of the standard linear model, $\mathbf{Z}_{ij}$ does not exist

We can write this equation in matrix-vector notation, the same way we did in Equation 2.2, but only for one cluster at a time.

$$\mathbf{Y}_j = \mathbf{X}_j\boldsymbol{\gamma} + \mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\varepsilon}_j \qquad j \in \{1,...,J\}$$

- $\mathbf{Y}_j$ is an $(n_j \times 1)$ response vector
- $\mathbf{X}_j$ is an $(n_j \times Q)$ model matrix for the fixed effects
- $\boldsymbol{\gamma}$ is an $(Q \times 1)$ vector of the fixed-effect coefficients
- $\mathbf{Z}_j$ is an $(n_j \times P)$ model matrix for the random effects
- $\mathbf{u}_j$ is an $(P \times 1)$ vector of random-effects coefficients. We assume $\mathbf{u}_j$ is normally distributed with multivariate normal distribution $\mathbf{u}_j \sim N_P(0, \mathbf{T})$
- $\boldsymbol{\varepsilon}_j$ is an $(n_j \times 1)$ vector of errors. We assume $\boldsymbol{\varepsilon}_j$ is normally distributed with multivariate normal distribution $\boldsymbol{\varepsilon}_j \sim N_{n_j}(0, \sigma^2\boldsymbol{\Lambda}_j)$
  - In most studies the error does not variate between clusters, in that case $\boldsymbol{\Lambda}_j = \mathbf{I}_{n_j}$
  - We assume that all $\boldsymbol{\varepsilon}_j$ and $\boldsymbol{\varepsilon}_{j'}$ are independent for $j \neq j'$
  - We assume that $Cov(\boldsymbol{\varepsilon}_j, \mathbf{u}_j) = 0$ for all j

Combining all clusters, the matrix-vector equation is the following:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \tag{2.7}$$

- $\mathbf{Y}$ is an $(N \times 1)$ response vector $([\mathbf{Y}_1^T, ..., \mathbf{Y}_J^T]^T)$
- $\mathbf{X}$ is an $(N \times Q)$ model matrix for the fixed effects $([X_1^T, ..., X_J^T]^T)$
- $\boldsymbol{\gamma}$ is an $(Q \times 1)$ vector of fixed-effect coefficients
- $\mathbf{Z}$ is an $(N \times J \cdot P)$ model matrix for the random effects.

- $\mathbf{u}$ is an $(J \cdot P \times 1)$ vector of random-effects coefficients $([\mathbf{u}_1^T, ..., \mathbf{u}_J^T]^T)$
  - We assume $\mathbf{u}$ is normally distributed, with normal distribution $\mathbf{u} \sim N_{J.P}(0, \mathbf{T}^*)$

$$\mathbf{T}^* = \begin{bmatrix} \mathbf{T} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{T} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{T} \end{bmatrix}$$

- $\boldsymbol{\varepsilon}$ is an $(N \times 1)$ vector of errors $([\boldsymbol{\varepsilon}_1^T, ..., \boldsymbol{\varepsilon}_J^T]^T)$
  - We assume $\boldsymbol{\varepsilon}$ is normally distributed with normal distribution $\boldsymbol{\varepsilon} \sim N_N(0, \sigma^2 \boldsymbol{\Lambda})$

$$\sigma^2 \boldsymbol{\Lambda} = \sigma^2 \begin{bmatrix} \boldsymbol{\Lambda}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Lambda}_J \end{bmatrix}$$

2.2.1. Estimating parameters

When the data are gathered, we need a good estimation of the regression coefficients $\boldsymbol{\gamma}$ and $\mathbf{u}$. In this subsection the method for obtaining the estimates of those parameters are explained.
The estimates for both parameters rely on the variance of the model. Therefore we need an expression for both:

$$\mathrm{E}(\mathbf{Y}) = \mathrm{E}(\mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}) = \mathrm{E}(\mathbf{X}\boldsymbol{\gamma}) + \mathrm{E}(\mathbf{Z}\mathbf{u}) + \mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{X}\boldsymbol{\gamma} + 0 + 0 = \mathbf{X}\boldsymbol{\gamma}$$

$$\boldsymbol{\Theta} = \mathrm{Var}(\mathbf{Y}) = \mathrm{Var}(\mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}) \overset{independence}{=} \mathrm{Var}(\mathbf{X}\boldsymbol{\gamma}) + \mathrm{Var}(\mathbf{Z}\mathbf{u}) + \mathrm{Var}(\boldsymbol{\varepsilon}) = 0 + \mathbf{Z}\mathbf{T}^*\mathbf{Z}^T + \sigma^2\boldsymbol{\Lambda} = \mathbf{Z}\mathbf{T}^*\mathbf{Z}^T + \sigma^2\boldsymbol{\Lambda}$$

When both $\boldsymbol{\gamma}$ and $\mathbf{u}$ are estimated, it will give the following expression:

Lemma 2.2.1. *Let $\mathbf{Y} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$ be a linear mixed-effects regression model, with $\boldsymbol{\Theta}$ the variance of the model. The estimators for $\boldsymbol{\gamma}$ and $\mathbf{u}$ are given by:*

$$\begin{aligned} \hat{\boldsymbol{\gamma}} &= (\mathbf{X}^T\boldsymbol{\Theta}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Theta}^{-1}\mathbf{Y} \\ \hat{\mathbf{u}} &= \mathbf{T}^*\mathbf{Z}^T\boldsymbol{\Theta}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\gamma}}) \end{aligned} \tag{2.8}$$

Proof. *(Based on Fox, 2016)*
To begin the proof of this Lemma, we need the estimating equation. This equation can be derived by the Generalized Least Squares (GLS) criterion of the mixed-effects model.

$$Q(\boldsymbol{\gamma}, \mathbf{u}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\gamma} - \mathbf{Z}\mathbf{u})^T\boldsymbol{\Lambda}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\gamma} - \mathbf{Z}\mathbf{u}) + \sigma^2\mathbf{u}^T\mathbf{T}^{*-1}\mathbf{u}$$

We can find the derivative of this equation with respect to both $\boldsymbol{\gamma}$ and $\mathbf{u}$ by using the identities of the scalar-by-vector derivative ("Matrix calculus", 2025), and the fact that $\boldsymbol{\Lambda}$ and $\mathbf{T}^*$ and thus $\boldsymbol{\Lambda}^{-1}$ and $\mathbf{T}^{*-1}$ are symmetric.

$$\begin{aligned} \frac{\partial Q}{\partial \boldsymbol{\gamma}} &= -2\mathbf{X}^T\boldsymbol{\Lambda}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\gamma} - \mathbf{Z}\mathbf{u}) \\ \frac{\partial Q}{\partial \mathbf{u}} &= -2\mathbf{Z}^T\boldsymbol{\Lambda}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\gamma} - \mathbf{Z}\mathbf{u}) + 2\sigma^2\mathbf{T}^{*-1}\mathbf{u} \end{aligned}$$

To obtain the minimal error while estimating the parameters, the derivative has to be equal to 0, setting the equations to 0 will give:

$$\begin{aligned} \mathbf{X}^T\boldsymbol{\Lambda}^{-1}\mathbf{X}\boldsymbol{\gamma} + \mathbf{X}^T\boldsymbol{\Lambda}^{-1}\mathbf{Z}\mathbf{u} &= \mathbf{X}^T\boldsymbol{\Lambda}^{-1}\mathbf{Y} \\ \mathbf{Z}^T\boldsymbol{\Lambda}^{-1}\mathbf{X}\boldsymbol{\gamma} + (\mathbf{Z}^T\boldsymbol{\Lambda}^{-1}\mathbf{Z} + \sigma^2\mathbf{T}^{*-1})\mathbf{u} &= \mathbf{Z}^T\boldsymbol{\Lambda}^{-1}\mathbf{Y} \end{aligned}$$

When we write this in matrix-vector notation, we can obtain both estimators by solving the system. Which gives us Equation 2.8.

$$\begin{bmatrix} \mathbf{X}^T\boldsymbol{\Lambda}^{-1}\mathbf{X} & \mathbf{X}^T\boldsymbol{\Lambda}^{-1}\mathbf{Z} \\ \mathbf{Z}^T\boldsymbol{\Lambda}^{-1}\mathbf{X} & \mathbf{Z}^T\boldsymbol{\Lambda}^{-1}\mathbf{Z} + \sigma^2\mathbf{T}^{*-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\boldsymbol{\Lambda}^{-1}\mathbf{Y} \\ \mathbf{Z}^T\boldsymbol{\Lambda}^{-1}\mathbf{Y} \end{bmatrix}$$

□

2.3. Three-Level Linear Mixed Model

In Section 2.2 we looked into two-level linear mixed models, for example, students in a class or teachers in a school. But sometimes we need more levels, for example, in the case of students in a class, where some of them have the same teacher, or students in a class in the same school. In the three-level case, we do not only have a random effect on class-level, but also on teacher-level or school-level.
Equation 2.3 is extended to obtain an equation for the three-level model.

$$\begin{aligned} y_i &= \beta_{1jk} + \beta_{2jk}X_{2i} + \cdots + \beta_{Qjk}X_{Qi} + \varepsilon_{ijk} \qquad i \in \{1,\ldots,N\}, j \in \{1,\ldots,J\}, k \in \{1,\ldots,K\} \\ &= \boldsymbol{\beta}_{jk}\mathbf{X}_i + \varepsilon_{ijk} \end{aligned} \tag{2.9}$$

- y_i is the response variable for a certain data point i
- β_{qjk} is the slope coefficient for $(q \in \{1,\ldots,Q\})$. They form the $(1 \times Q)$ vector $\boldsymbol{\beta}_{jk}$
- X_{qi} is the explanatory variable for a data point i for $(q \in \{1,\ldots,Q\})$. They form the $(Q \times 1)$ vector $\mathbf{X}_i$
- ε_{ijk} is the measurement error of data point i. We assume ε_{ijk} is normally distributed with normal distribution $\varepsilon_{ijk} \sim N(0, \sigma^2_{jk})$

In the same way we added another level in Section 2.2, we can add one to the three-level model.

$$\beta_{qjk} = \pi_{qk} + r_{qjk} \qquad q \in \{1,\ldots,Q\}, j \in \{1,\ldots,J\}, k \in \{1,\ldots,K\} \tag{2.10}$$

- β_{qjk} is the regression coefficient for the qth explanatory variable, the jth second level cluster and the kth third level cluster
- π_{qk} is the fixed effect of the qth slope, independent of the second level cluster
- r_{qjk} is the random effect of the qth slope and the jth second level cluster. We assume r_{qjk} is normally distributed with normal distribution $r_{qjk} \sim N(0, \tau_{2q})$

We now have the same two levels as in Equation 2.4, but with slightly different notation. To add a third level, we use the fact that π_{qk} can also be split into a fixed and random effect.

$$\pi_{qk} = \gamma_q + u_{qk} \qquad q \in \{1,\ldots,Q\}, k \in \{1,\ldots,K\} \tag{2.11}$$

- π_{qk} is the fixed effect of the qth slope, independent of the second level cluster
- γ_q is the fixed effect of the qth slope, independent of the third level cluster
- u_{qk} is the random effect of the qth slope. We assume u_{qk} is normally distributed with normal distribution $u_{qk} \sim N(0, \tau_{3q})$

We can combine the equations for the three-level model (2.9), the regression coefficients (2.10) and the fixed effects (2.11), to get the following equation:

$$\begin{aligned} y_i &= (\gamma_1 + u_{1k} + r_{1jk}) + (\gamma_2 + u_{2k} + r_{2jk})X_{2i} + \cdots + (\gamma_Q + u_{Qk} + r_{Qjk})X_{Qi} + \varepsilon_{ijk} \\ &= (\boldsymbol{\gamma} + \mathbf{u}_k + \mathbf{r}_{jk})\mathbf{X}_i + \varepsilon_{ijk} \qquad i \in \{1,\ldots,N\}, j \in \{1,\ldots,J\}, k \in \{1,\ldots,K\} \end{aligned}$$

- $\boldsymbol{\gamma}$ is an $(1 \times Q)$ vector of fixed effects
- $\mathbf{u}_k$ is an $(1 \times Q)$ vector of random effects for the third level clusters
- $\mathbf{r}_{jk}$ is an $(1 \times Q)$ vector of random effects for the second level clusters

We can also write this equation in the Laird-Ware form (Laird and Ware, 1982); we will split the equation into three components: the fixed component, the random component for the third level clusters and the random component for the second level clusters.

$$\begin{aligned} y_i &= \gamma_1 + \gamma_2 X_{2i} + \cdots + \gamma_Q X_{Qi} + u_{1k} + u_{2k}X_{2i} + \cdots + u_{Pk}X_{Pi} + r_{1jk} + r_{2jk}X_{2i} + \cdots + r_{Rjk}X_{Ri} + \varepsilon_{ijk} \\ &= \boldsymbol{\gamma}\mathbf{X}_i + \mathbf{u}_k\mathbf{Z}_i^{(3)} + \mathbf{r}_{jk}\mathbf{Z}_i^{(2)} + \varepsilon_{ijk} \qquad i \in \{1,\ldots,N\}, j \in \{1,\ldots,J\}, k \in \{1,\ldots,K\} \end{aligned}$$

- $\mathbf{Z}_i^{(3)}$ is an $(P \times 1)$ vector which contains the explanatory variables only if they are modeled with a random effect in the third level
  - In some cases $\mathbf{X}_i = \mathbf{Z}_i^{(3)}$, when every slope has a random effect in the third level
- $\mathbf{Z}_i^{(2)}$ is an $(R \times 1)$ vector which contains the explanatory variables only if they are modeled with a random effect in the second level
  - In some cases $\mathbf{X}_i = \mathbf{Z}_i^{(2)}$, when every slope has a random effect in the third level

We can write this equation in matrix-vector notation, the same way we did in Equation 2.7.

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}^{(3)}\mathbf{u} + \mathbf{Z}^{(2)}\mathbf{r} + \boldsymbol{\varepsilon}$$

- $\mathbf{Y}$ is an $(N \times 1)$ response vector
- $\mathbf{X}$ is an $(N \times Q)$ model matrix for the fixed effects
- $\boldsymbol{\gamma}$ is an $(Q \times 1)$ vector of fixed-effect coefficients
- $\mathbf{Z}^{(3)}$ is an $(N \times K \cdot P)$ model matrix for the random effects in the third level
- $\mathbf{u}$ is an $(K \cdot P \times 1)$ vector of random effects coefficients for the third level
  - We assume $\mathbf{u}$ is normally distributed with normal distribution $\mathbf{u} \sim N_{K \cdot P}(0, \mathbf{T}_3)$
- $\mathbf{Z}^{(2)}$ is an $(N \times J \cdot R)$ model matrix for the random effects in the second level
- $\mathbf{r}$ is an $(J \cdot R \times 1)$ vector of random effects coefficients for the second level
  - We assume $\mathbf{r}$ is normally distributed with normal distribution $\mathbf{r} \sim N_{J \cdot R}(0, \mathbf{T}_2)$
- $\boldsymbol{\varepsilon}$ is an $(N \times 1)$ vector of errors
  - We assume $\boldsymbol{\varepsilon}$ is normally distributed with normal distribution $\boldsymbol{\varepsilon} \sim N_N(0, \sigma^2\boldsymbol{\Lambda})$

2.3.1. Estimating parameters

When the data are gathered, we need a good estimation of the regression coefficients $\boldsymbol{\gamma}$, $\mathbf{u}$ and $\mathbf{r}$. In this subsection the method for obtaining the estimates of those parameters are explained.
The estimates for all parameters rely on the variance of the model. Therefore we need an expression for both:

$$\begin{aligned}
\mathrm{E}(\mathbf{Y}) &= \mathrm{E}(\mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}^{(3)}\mathbf{u} + \mathbf{Z}^{(2)}\mathbf{r} + \boldsymbol{\varepsilon}) = \mathrm{E}(\mathbf{X}\boldsymbol{\gamma}) + \mathrm{E}(\mathbf{Z}^{(3)}\mathbf{u}) + \mathrm{E}(\mathbf{Z}^{(2)}\mathbf{r}) + \mathrm{E}(\boldsymbol{\varepsilon}) \\
&= \mathbf{X}\boldsymbol{\gamma} + 0 + 0 + 0 = \mathbf{X}\boldsymbol{\gamma} \\
\boldsymbol{\Theta} = \mathrm{Var} &= \mathrm{Var}(\mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}^{(3)}\mathbf{u} + \mathbf{Z}^{(2)}\mathbf{r} + \boldsymbol{\varepsilon}) \overset{independence}{=} \mathrm{Var}(\mathbf{X}\boldsymbol{\gamma}) + \mathrm{Var}(\mathbf{Z}^{(3)}\mathbf{u}) + \mathrm{Var}(\mathbf{Z}^{(2)}\mathbf{r}) + \mathrm{Var}(\boldsymbol{\varepsilon}) \\
&= 0 + \mathbf{Z}^{(3)}\mathbf{T}_3\mathbf{Z}^{(3)T} + \mathbf{Z}^{(2)}\mathbf{T}_2\mathbf{Z}^{(2)T} + \sigma^2\boldsymbol{\Lambda} = \mathbf{Z}^{(3)}\mathbf{T}_3\mathbf{Z}^{(3)T} + \mathbf{Z}^{(2)}\mathbf{T}_2\mathbf{Z}^{(2)T} + \sigma^2\boldsymbol{\Lambda}
\end{aligned}$$

We can estimate the unknown coefficients and get the following result:

Lemma 2.3.1. *Let* $\mathbf{Y} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}^{(3)}\mathbf{u} + \mathbf{Z}^{(2)}\mathbf{r} + \boldsymbol{\varepsilon}$ *be a linear mixed-effects regression model, with* $\boldsymbol{\Theta}$ *the variance of the model. The estimators for* $\boldsymbol{\gamma}$, $\mathbf{u}$ *and* $\mathbf{r}$ *are given by:*

$$\begin{aligned}
\hat{\boldsymbol{\gamma}} &= (\mathbf{X}^T\boldsymbol{\Theta}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Theta}^{-1}\mathbf{Y} \\
\hat{\mathbf{u}} &= \mathbf{T}_3\mathbf{Z}^{(3)T}\boldsymbol{\Theta}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\gamma}}) \\
\hat{\mathbf{r}} &= \mathbf{T}_2\mathbf{Z}^{(2)T}\boldsymbol{\Theta}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\gamma}})
\end{aligned} \tag{2.12}$$

Proof. (based on Fox, 2016)
For the proof of this lemma, we need the estimating equation. This equation can be derived by the GLS criterion of the mixed-effects model.

$$Q(\boldsymbol{\gamma}, \mathbf{u}, \mathbf{r}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\gamma} - \mathbf{Z}^{(3)}\mathbf{u} - \mathbf{Z}^{(2)}\mathbf{r})^T\boldsymbol{\Lambda}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\gamma} - \mathbf{Z}^{(3)}\mathbf{u} - \mathbf{Z}^{(2)}\mathbf{r}) + \sigma^2\mathbf{u}^T\mathbf{T}_3^{-1}\mathbf{u} + \sigma^2\mathbf{r}^T\mathbf{T}_2^{-1}\mathbf{r}$$

We can take the derivative with respect to the coefficients by using the properties of matrix-vector derivative ("Matrix calculus", 2025), and using that $\mathbf{\Lambda}$, $\mathbf{T}_2$ and $\mathbf{T}_3$ and therefore $\mathbf{\Lambda}^{-1}$, $\mathbf{T}_2^{-1}$ and $\mathbf{T}_3^{-1}$ are symmetric.

$$\frac{\partial Q}{\partial \boldsymbol{\gamma}} = -2\mathbf{X}^T\mathbf{\Lambda}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\gamma}-\mathbf{Z}^{(3)}\mathbf{u}-\mathbf{Z}^{(2)}\mathbf{r})$$
$$\frac{\partial Q}{\partial \mathbf{u}} = -2\mathbf{Z}^{(3)T}\mathbf{\Lambda}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\gamma}-\mathbf{Z}^{(3)}\mathbf{u}-\mathbf{Z}^{(2)}\mathbf{r})+2\sigma^2\mathbf{T}_3^{-1}\mathbf{u}$$
$$\frac{\partial Q}{\partial \mathbf{r}} = -2\mathbf{Z}^{(2)T}\mathbf{\Lambda}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\gamma}-\mathbf{Z}^{(3)}\mathbf{u}-\mathbf{Z}^{(2)}\mathbf{r})+2\sigma^2\mathbf{T}_2^{-1}\mathbf{r}$$

To obtain the minimal error while estimating the parameters, the derivative has to be equal to 0. Setting the derivatives equal to 0 will give:

$$\mathbf{X}^T\mathbf{\Lambda}^{-1}\mathbf{X}\boldsymbol{\gamma}+\mathbf{X}^T\mathbf{\Lambda}^{-1}\mathbf{Z}^{(3)}\mathbf{u}+\mathbf{X}^T\mathbf{\Lambda}^{-1}\mathbf{Z}^{(2)}\mathbf{r}=\mathbf{X}^T\mathbf{\Lambda}^{-1}\mathbf{Y}$$
$$\mathbf{Z}^{(3)}\mathbf{\Lambda}^{-1}\mathbf{X}\boldsymbol{\gamma}+\mathbf{Z}^{(3)}\mathbf{\Lambda}^{-1}\mathbf{Z}^{(2)}\mathbf{r}+(\mathbf{Z}^{(3)}\mathbf{\Lambda}^{-1}\mathbf{Z}^{(3)}+\sigma^2\mathbf{T}_3^{-1})\mathbf{u}=\mathbf{Z}^{(3)T}\mathbf{\Lambda}^{-1}\mathbf{Y}$$
$$\mathbf{Z}^{(2)}\mathbf{\Lambda}^{-1}\mathbf{X}\boldsymbol{\gamma}+\mathbf{Z}^{(2)}\mathbf{\Lambda}^{-1}\mathbf{Z}^{(3)}\mathbf{u}+(\mathbf{Z}^{(2)}\mathbf{\Lambda}^{-1}\mathbf{Z}^{(2)}+\sigma^2\mathbf{T}_2^{-1})\mathbf{r}=\mathbf{Z}^{(2)T}\mathbf{\Lambda}^{-1}\mathbf{Y}$$

When we write this in matrix-vector notation, we can obtain the estimators by solving the system. Which gives us Equation 2.12.

$$\begin{bmatrix} \mathbf{X}^T\mathbf{\Lambda}^{-1}\mathbf{X} & \mathbf{X}^T\mathbf{\Lambda}^{-1}\mathbf{Z}^{(3)} & \mathbf{X}^T\mathbf{\Lambda}^{-1}\mathbf{Z}^{(2)} \\ \mathbf{Z}^{(3)}\mathbf{\Lambda}^{-1}\mathbf{X} & \mathbf{Z}^{(3)}\mathbf{\Lambda}^{-1}\mathbf{Z}^{(3)}+\sigma^2\mathbf{T}_3^{-1} & \mathbf{Z}^{(3)}\mathbf{\Lambda}^{-1}\mathbf{Z}^{(2)} \\ \mathbf{Z}^{(2)}\mathbf{\Lambda}^{-1}\mathbf{X} & \mathbf{Z}^{(2)}\mathbf{\Lambda}^{-1}\mathbf{Z}^{(3)} & \mathbf{Z}^{(2)}\mathbf{\Lambda}^{-1}\mathbf{Z}^{(2)}+\sigma^2\mathbf{T}_2^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma} \\ \mathbf{u} \\ \mathbf{r} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{\Lambda}^{-1}\mathbf{Y} \\ \mathbf{Z}^{(3)T}\mathbf{\Lambda}^{-1}\mathbf{Y} \\ \mathbf{Z}^{(2)T}\mathbf{\Lambda}^{-1}\mathbf{Y} \end{bmatrix}$$

□

2.4. Comparison Values

After the models are fitted on the collected data, we have to find out if a model is better or worse than another model. There are a lot of different values for testing if a model is good or bad. The ones that will be used in this research are R^2, AIC and BIC. In this section, how these values are computed and what they tell about the model will be discussed.

2.4.1. R-squared

R^2 is the most used value for comparing models. It is the squared multiple correlation, and it represents the proportion of variance in the response variable captured by the regression (Fox, 2016). In other words, R^2 represents how much of the error ε is explained by the regression model.
In every model R^2 is constructed differently. How R^2 is computed in all the used models will be explained in this section.

Linear Models

The most simple version of R-squared is the one used in the linear regression model. We divide the variance in our regression model by the total variance in the model; in that way, we obtain a value for the amount of variance that is explained by the model. (Fox, 2016).

$$R^2 = \frac{SS_{reg}}{SS_{tot}}$$

- SS_{reg} is the regression sum of squares, the variance in the model.

$$SS_{reg} = \sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2$$

  - $\bar{y}$ is the mean observation
  - $\hat{y}_i$ is the regressed ith observation

- SS_{tot} is the total sum of squares, the total variance in the observations.

$$SS_{tot} = \sum_{i=1}^{N}(y_i - \bar{y})^2$$

- y_i is the ith observation
- $\bar{y}$ is the mean observation

We can also use the adjusted R^2, which penalizes the value of R^2 for the degrees of freedom, so if more variables are added, we have fewer degrees of freedom, so we should have a smaller adjusted R^2. The equation for the adjusted R^2 is given by (Fox, 2016):

$$\bar{R}^2 = 1 - \frac{\frac{SS_{res}}{N-Q}}{\frac{SS_{tot}}{N-1}}$$

- SS_{res} is the residual sum of squares, the variance in the residuals.

$$SS_{res} = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

  - $\hat{y}_i$ is the regressed ith observation
  - y_i is the ith observation

- SS_{tot} is the total sum of squares, the total variance in the observations.

$$SS_{tot} = \sum_{i=1}^{N}(y_i - \bar{y})^2$$

  - y_i is the ith observation
  - $\bar{y}$ is the mean observation

Linear two-level Mixed Models
In the linear mixed-effect model, we cannot use the values for SS_{res} and SS_{tot}, because we do not only have variance in the residuals, but also in the levels. We still divide the variance in the model by the total variance. The variance in the model is the variance in the fixed effects, and the total variance is the variance in the observations, the fixed effects, and the random effects.
We also distinguish two different kinds of R-squared: R_m^2 and R_c^2, the explained variance in the fixed effects and the explained variance in the random effects (Rights and Sterba, 2023)

$$\begin{aligned} R_m^2 &= \frac{\sigma_f^2}{\sigma_f^2 + \tau^2 + \sigma^2} \\ R_c^2 &= \frac{\sigma_f^2 + \tau^2}{\sigma_f^2 + \tau^2 + \sigma^2} \end{aligned} \tag{2.13}$$

- σ_f^2 is the variance in the fixed effects only of the model

$$\hat{\sigma}_f^2 = \mathrm{Var}(\mathbf{X}\boldsymbol{\gamma}) = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{\gamma}\mathbf{X}_i - \bar{\boldsymbol{\gamma}\mathbf{X}})^2$$

- τ^2 is the in-between variance in the model
- σ^2 is the residual variance in the model

To obtain R^2, we need an estimator for both σ^2 and τ^2. $\hat{\sigma}^2$ is obtained by using restricted maximum likelihood estimation. Which gives the following result (Finch et al., 2014):

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{J}(n_j - 1)S_j^2}{N-J} \tag{2.14}$$

- S_j^2 is the variance within a cluster j

$$\hat{S}_j^2 = \frac{\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)}{(n_j - 1)}$$

- y_{ij} is the ith observation in the jth cluster
- $\bar{y}_j$ is the mean observation in the jth cluster

After σ^2 is estimated, we can obtain an estimator for τ^2, given by (Finch et al., 2014):

$$\hat{\tau}^2 = S_B^2 - \frac{\hat{\sigma}^2}{\tilde{n}} \tag{2.15}$$

- S_B^2 is the weighted in between cluster variance

$$\hat{S}_B^2 = \frac{\sum_{j=1}^{J} n_j(\bar{y}_j - \bar{y})^2}{\tilde{n}(J-1)}$$

  - $\bar{y}_j$ is the mean observation in the jth cluster
  - $\bar{y}$ is the overall mean of the observations

- $$\tilde{n} = \frac{1}{J-1}(N - \frac{\sum_{j=1}^{J} n_j^2}{N})$$

Linear three-level Mixed Models
In the three-level model, we also have variance in the third level, which will extend Equation 2.13 for R_m^2 and R_c^2 to (Rights and Sterba, 2023):

$$R_m^2 = \frac{\sigma_f^2}{\sigma_f^2 + \tau_2^2 + \tau_3^2 + \sigma^2}$$
$$R_c^2 = \frac{\sigma_f^2 + \tau_2^2 + \tau_3^2}{\sigma_f^2 + \tau_\beta^2 + \tau_\pi^2 + \sigma^2}$$

- σ_f^2 is the variance in the fixed effects only of the model

$$\hat{\sigma}_f^2 = \text{Var}(\mathbf{X}\boldsymbol{\gamma}) = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{\gamma}\mathbf{X}_i - \bar{\boldsymbol{\gamma}\mathbf{X}})^2$$

- τ_2^2 is the in-between variance in the second level of the model
- τ_3^2 is the in-between variance in the third level of the model
- σ^2 is the residual variance in the model

Because we have a new equation, the estimators for σ^2, τ_2^2 and τ_3^2 needed to be recalculated. τ_2^2 and τ_3^2 are calculated similarly to Equation 2.15. σ^2 is also calculated with the restricted maximum likelihood method, but in contrast to Equation 2.14, we cannot give an explicit expression for $\hat{\sigma}^2$. The log-likelihood equation is given by (Maestrini et al., 2025):

$$\ell_R(\boldsymbol{\gamma}, \mathbf{Y}, \mathbf{u}, \mathbf{r}) = -\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\gamma} - \mathbf{Z}^{(3)}\mathbf{u} - \mathbf{Z}^{(2)}\mathbf{r})^T \boldsymbol{\Lambda}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\gamma} - \mathbf{Z}^{(3)}\mathbf{u} - \mathbf{Z}^{(2)}\mathbf{r})$$

If we take the derivative with respect to σ^2 and set it equal to 0, we will get the estimator $\hat{\sigma}^2$.

As mentioned earlier, the expression for τ_2^2 and τ_3^2 are similar to the expression for the 2-level model (Snijders and Bosker, 2011):

$$\hat{\tau}_2^2 = S_2^2 - \frac{\hat{\sigma}^2}{\tilde{n}_2}$$
$$\hat{\tau}_3^2 = S_3^2 - \frac{\hat{\tau}_2^2\tilde{n}_2 + \hat{\sigma}^2}{\tilde{n}_3}$$

- S_2^2 is the in-between cluster variance in the second level

$$\hat{S}_2^2 = \frac{\sum_{j=1}^{J} n_j(\bar{y}_j - \bar{y})^2}{\tilde{n}_2(J-1)}$$

  - $\bar{y}_j$ is the mean y in the jth cluster
  - $\bar{y}$ is the mean y in the whole sample

- $$\tilde{n}_2 = \frac{1}{J-1}(N - \frac{\sum_{j=1}^{J} n_j^2}{N})$$

- S_3^2 is the in-between cluster variance in the second level

$$S_3^2 = \frac{\sum_{k=1}^{K} n_k(\bar{y}_k - \bar{y})^2}{\tilde{n}_3(K-1)}$$

  - $\bar{y}_k$ is the mean y in the kth cluster
  - $\bar{y}$ is the mean y in the whole sample

- $$\tilde{n}_3 = \frac{1}{K-1}(N - \frac{\sum_{k=1}^{K} n_k^2}{N})$$

2.4.2. AIC

The coefficient of determination, R^2, measures the proportion of variance in the dependent variable explained by the model. However, R^2 never decreases when additional predictors are included in a model, even when those predictors are not significant. As a result, R^2 alone tends to choose complex models over simple models. To address this limitation, information criteria such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are used for model selection. Both criteria balance model fit and model complexity by penalizing the inclusion of additional parameters (Sohil et al., 2022). Lower AIC and BIC values indicate a better balance between goodness of fit and model selection. So, AIC and BIC are used to obtain a value for the fitness of the model. A smaller value is associated with a better fit.

The equation for AIC is given below; the part in this equation, $2Q$, will give us a bigger AIC if we use more explanatory variables. Since the lower AIC, the better the model, a penalty for more predictors is included. (Greven and Kneib, 2010)

$$AIC = -2log(L) + 2Q$$

- L the likelihood function of the model

2.4.3. BIC

BIC is similar to AIC but imposes a stronger penalty for model complexity. The penalty term depends on the samples size (N); therefore, as the sample size increases, BIC increasingly favors models with a small sample size compared with AIC. The part in the equation below, $Qlog(N)$, is now extended with a $log(N)$. So, bigger values of N result in bigger values of BIC compared to AIC (Shen and González, 2021).

$$BIC = -2log(L) + Qlog(N)$$

- L is the likelihood function of the model

3

Method

To answer the research question, which model best explains the influence of math teachers on students' math anxiety, multiple models were compared. This chapter first discusses the data collected in the pilot study (Section 3.1). Subsequently, Section 3.2 contains the construction of variables, using the collected data, and how they will be used in the research models. In the last section (3.3), the construction of the research models for the data sample is discussed.

3.1. Sample data

For this pilot study, data were collected at a Dutch high school. The sample consisted of six teachers, fourteen classes, and 93 students. Further details regarding the participating teachers, classes, and students are provided in Appendix B.
Teachers completed a questionnaire concerning instructional objectives, teaching-related emotions, and mindset. The questions can be found in Appendix A.1. The students were given a questionnaire about their math anxiety and their perception of their teachers' interpersonal behavior. The student questionnaire can be found in Appendix A.2.

The first section of the teacher questionnaire consisted of seven items about the mathematical instructional objectives. The questionnaire was developed and validated by Lee (2025). The seven items measured the extent to which the instruction of each teacher was concept-focused or procedure-focused. Teachers had to state how much emphasis they placed on certain aspects of mathematical instruction. Teachers responded on a four-point scale, ranging from no emphasis to heavy emphasis. The first five items were concept-focused and the last two items were procedure-focused.
Concept-focused instruction emphasizes connections between mathematical ideas and promotes conceptual understanding. An example of concept-focused instruction is building connections between different mathematical concepts such as fractions and division. Procedure-focused instruction emphasizes the application of algorithms and procedures, such as computing derivatives, rather than understanding the underlying concepts.

The next section of the questionnaire assessed teaching-related emotions: enjoyment, anxiety, and anger. The questionnaire was developed by Frenzel (2016). Each emotion was measured using four items. Teachers had to state to what extent they agreed with the items. Responses were given on a four-point scale ranging from strongly disagree to strongly agree.
The items were administered twice for each emotion. Once for teaching in general, in this case, questions were of the format "I generally enjoy teaching." The second time these items were asked, respondents were instructed to consider a specific class they taught, for example, the class they taught on Monday morning. So, the items were phrased as "I enjoy teaching these students".

The final section assessed teachers' mindset, specifically their growth and fixed mindset beliefs. Teachers' growth mindset was measured using six items (Shoshani, 2021). The first four items and the final item reflected a fixed mindset, while the second-to-last item was focused on a growth mindset. Teachers had to

indicate the extent to which they agreed with the statements. They could choose between six options ranging from totally disagree to totally agree.
A fixed mindset reflects the belief that intelligence and mathematical ability are largely innate and difficult to change. In contrast, a growth mindset reflects the belief that mathematical ability can be developed through effort and learning.

The student questionnaire began with nine items measuring their math anxiety (Schmitz et al., 2022). Each item described a mathematical situation, for example, thinking about a math test or using the tables in a mathematics book. The students had to rate the level of anxiety they experienced in those situations. Responses were given on a five-point scale ranging from low anxiety to high anxiety.

The final section measured students' perceptions of teacher-student interpersonal behavior. This questionnaire was the abbreviated version of the questionnaire based on Leary's Rose (Wubbels et al., 2024). The instrument distinguishes eight dimensions: leadership, helping/friendly, understanding, student responsibility/freedom, uncertain, dissatisfied, admonishing and strict. Every student had to answer three items about each dimension. They had to indicate to what extent they agreed with each statement, using a five-point response scale, ranging from strongly disagree up to strongly agree.

3.2. Variable construction

The questionnaire responses were aggregated into several variables that were used in the statistical models. This section describes how each variable was computed.

The mathematical instructional objectives scale consisted of seven items. They were denoted by IO1 to IO7. Each item was assigned a value based on the teacher's response, ranging from 1 (no emphasis) to 4 (heavy emphasis). The concept-focused instruction score (CFI) was computed as the mean of the items IO1 to IO5. The procedure-focused instruction score (PFI) was computed as the mean of items IO6 and IO7.

The teaching-related emotions questionnaire consisted of 24 items. The enjoyment items were denoted by EG1-EG4 for general enjoyment and EC1-EC4 for class-specific enjoyment. Analogous notation was used for anxiety (AX) and anger (AG). Each item was scored on a four-point Likert scale ranging from 1 (strongly disagree) to 4 (strongly agree). Mean scores were calculated for EG, EC, AXG, AXC, AGG, and AGC, representing general enjoyment, class-specific enjoyment, general anxiety, class-specific anxiety, general anger, and class-specific anger, respectively.

The mindset scale consisted of six items, denoted by GM1 to GM6. The mean score across these six items was computed to obtain an overall mindset score (GM). Lower scores of GM indicated a more fixed mindset, whereas higher scores indicated a stronger growth mindset.

The math anxiety scale consisted of nine items, denoted by MA1 to MA9, each scored on a 5-point Likert scale ranging from 1 (low anxiety) to 5 (high anxiety). The math anxiety score (MA) was computed as the mean of the items MA1 to MA9. Higher scores indicated higher levels of math anxiety.

The interpersonal behavior questionnaire measured eight dimensions: leadership, helping/friendly, understanding, student responsibility/freedom, uncertain, dissatisfied, admonishing, and strict. The eight dimensions were denoted by $X_1, \ldots, X_8$. Each dimension consisted of three items scored on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Dimension scores were computed as the mean of the corresponding three items.

In Table 3.1 an overview of the different variables and their characteristics can be found.

3.3. Research models

Using the data described in Section 3.1, three statistical models were constructed: the Teacher model, the Class model, and the Student model. The Teacher model (3.3.1) is a linear regression model, the Class model (3.3.2) is a two-level mixed-effects model, and the Student model (3.3.3) is a three-level mixed-effects model.

Questionnaire	Characteristic	Variable	Student/Teacher	Value
Instructional objectives	Procedure focused instruction	PFI	Teacher	continuous, [1,4]
	Concept focused instruction	CPI	Teacher	continuous, [1,4]
Teacher emotions	General Enjoyment	EG	Teacher	continuous, [1,4]
	General Anxiety	AXG	Teacher	continuous, [1,4]
	General Anger	AGG	Teacher	continuous, [1,4]
	Class Enjoyment	EC	Teacher	continuous, [1,4]
	Class Anxiety	AXC	Teacher	continuous, [1,4]
	Class Anger	AGC	Teacher	continuous, [1,4]
Growth Mindset	Growth mindset	GM	Teacher	continuous, [1,6]
Math Anxiety	Math Anxiety	MA	Student	continuous, [1,5]
Interpersonal behavior	Leadership	X_1	Student	continuous, [1,5]
	Helping/Friendly	X_2	Student	continuous, [1,5]
	Understanding	X_3	Student	continuous, [1,5]
	Student Responsibility/Freedom	X_4	Student	continuous, [1,5]
	Uncertain	X_5	Student	continuous, [1,5]
	Dissatisfied	X_6	Student	continuous, [1,5]
	Admonishing	X_7	Student	continuous, [1,5]
	Strict	X_8	Student	continuous, [1,5]

Table 3.1: Variables retrieved from the data sample

3.3.1. Teacher model

The first model operates at the teacher level. Each teacher completed a questionnaire measuring various aspects of teacher behavior, from which six explanatory variables were derived. The students answered questions about their perceptions of their teachers, from which eight variables were obtained. This model investigates the relationship between these teacher characteristics and the average math anxiety of the students taught by that teacher.
The regression equation for the Teacher model is given by:

$$\begin{aligned}\overline{MA}_j =& \beta_0 + \beta_1 PFI_j + \beta_2 CPI_j + \beta_3 EG_j + \beta_4 AXG_j + \beta_5 AGG_j + \beta_6 GM_j + \beta_7 \overline{X}_{1j} + \beta_8 \overline{X}_{2j} + \beta_9 \overline{X}_{3j} + \beta_9 \overline{X}_{4j} + \beta_{10} \overline{X}_{5j} \\ &+ \beta_{11} \overline{X}_{6j} + \beta_{12} \overline{X}_{7j} + \beta_{13} \overline{X}_{8j} + \varepsilon_j \qquad j \in \{1, \ldots, J\}\end{aligned} \tag{3.1}$$

- $\overline{MA}_j = \frac{1}{n_j} \sum_{k=1}^{n_j} MA_k \qquad j \in \{1, \ldots, J\}$
- $\overline{X}_{lj} = \frac{1}{n_j} \sum_{k=1}^{n_j} X_{lk} \qquad j \in \{1, \ldots, J\}, l \in \{1, \ldots, 8\}$
- We assume ε_j has a normal distribution $\varepsilon \sim N(0, \sigma^2)$
- Definitions of all variables are provided in Table 3.1

In matrix notation, the model can be written as:

$$\overline{\mathbf{MA}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{3.2}$$

- $\overline{\mathbf{MA}}$ is an $(J \times 1)$ vector of the average math anxiety per teacher
- $\boldsymbol{\beta}$ is an (15×1) vector of coefficients for the explanatory variables
- $\mathbf{X}$ is an $(J \times 15)$ matrix containing the explanatory variables
- $\boldsymbol{\varepsilon}$ is an $(J \times 1)$ vector of errors. We assume $\boldsymbol{\varepsilon}$ is normally distributed with normal distribution $\boldsymbol{\varepsilon} \sim N_J(0, \sigma^2 \mathbf{I})$

3.3.2. Class model

The second model operates at the class level. Each student completed a questionnaire measuring their teacher's interpersonal behavior, from which eight variables were obtained. This model investigates the relationship between these interpersonal characteristics and the average math anxiety of the students in a class. The regression equation for the Class model is given by:

$$\begin{aligned}\overline{MA}_k &= \beta_{j0} + \beta_{j1}\overline{X}_{1k} + \beta_{j2}\overline{X}_{2k} + \beta_{j3}\overline{X}_{3k} + \beta_{j4}\overline{X}_{4k} + \beta_{j5}\overline{X}_{5k} + \beta_{j6}\overline{X}_{6k} + \beta_{j7}\overline{X}_{7k} + \beta_{j8}\overline{X}_{8k} + \varepsilon_{jk} \\ &k \in \{1,...,K\}, j \in \{1,...,J\}\end{aligned} \tag{3.3}$$

- $\overline{MA}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} MA_i \qquad k \in \{1,...,K\}$
- $\overline{X}_{qk} = \frac{1}{n_k}\sum_{i=1}^{n_k} X_{qi} \qquad k \in \{1,...,K\}, q \in \{1,...,8\}$
- We assume ε_{jk} has a normal distribution $\varepsilon \sim N(0,\sigma^2)$
- Definitions of all variables are provided in Table 3.1

In matrix notation, the model can be written as:

$$\overline{\mathbf{MA}} = \mathbf{X}\boldsymbol{\beta}_j + \boldsymbol{\varepsilon} \qquad j \in \{1,...,J\} \tag{3.4}$$

- $\overline{\mathbf{MA}}$ is an $(K \times 1)$ vector of the average math anxiety per class
- $\boldsymbol{\beta}_j$ is an (8×1) vector of coefficients for the explanatory variables, they differ for each j
- $\mathbf{X}$ is an $(K \times 8)$ matrix containing the explanatory variables
- $\boldsymbol{\varepsilon}$ is an $(K \times 1)$ vector or errors. We assume $\boldsymbol{\varepsilon}$ is normally distributed with normal distribution $\boldsymbol{\varepsilon} \sim N_K(0,\sigma^2\mathbf{I})$

3.3.3. Student model

The first two models operated at the teacher level and the class level, so, the final model will operate at the student level. We will distinguish three cases: one model where students are grouped in clusters at a teacher- and class-level, a model where students are only clustered at class-level and one where students are clustered at teacher-level.

Three-level model

For the student model where students are grouped in two levels, we will dive deeper in the students' perception of their teacher. This model is a three-level mixed-effects model.

$$\begin{aligned}MA_i &= \beta_{0jk} + \beta_{1jk}X_{1i} + \beta_{2jk}X_{2i} + \beta_{3jk}X_{3i} + \beta_{4jk}X_{4i} + \beta_{5jk}X_{5i} + \beta_{6jk}X_{6i} + \beta_{7jk}X_{7i} + \beta_{8jk}X_{8i} + \varepsilon_{ijk} \\ &i \in \{1,...,N\}, j \in \{1,...,J\}, k \in \{1,...,K\}\end{aligned} \tag{3.5}$$

- We assume ε_{ijk} has a normal distribution $\varepsilon_{ijk} \sim N(0,\sigma^2)$
- Definitions of all variables are provided in Table 3.1

In matrix notation, the model can be written as:

$$\mathbf{MA} = \mathbf{X}\boldsymbol{\beta}_{jk} + \boldsymbol{\varepsilon} \qquad j \in \{1,...,J\}, k \in \{1,...,K\} \tag{3.6}$$

- $\mathbf{MA}$ is an $(N \times 1)$ vector of the math anxiety per student
- $\boldsymbol{\beta}_{jk}$ is an (8×1) vector of coefficients for the explanatory variables, they differ for each j and k
- $\mathbf{X}$ is an $(N \times 8)$ matrix containing the explanatory variables
- $\boldsymbol{\varepsilon}$ is an $(N \times 1)$ vector of errors. We assume $\boldsymbol{\varepsilon}$ is normally distributed with normal distribution $\boldsymbol{\varepsilon} \sim N(0,\sigma^2\mathbf{I})$

Student-Class model

The second sub-model takes a deeper dive into the effect of the interpersonal relationship of students within a class. This is a two-level mixed-effects model.

$$\begin{aligned} MA_i = {} & \beta_{0k} + \beta_{1k}X_{1i} + \beta_{2k}X_{2i} + \beta_{3k}X_{3i} + \beta_{4k}X_{4i} + \beta_{5k}X_{5i} + \beta_{6k}X_{6i} + \beta_{7k}X_{7i} + \beta_{8k}X_{8i} + \varepsilon_{ik} \\ & i \in \{1,\ldots,N\}, k \in \{1,\ldots,K\} \end{aligned} \tag{3.7}$$

- We assume ε_{ik} has a normal distribution $\varepsilon_{ik} \sim N(0,\sigma^2)$
- Definitions of all variables are provided in Table 3.1

In matrix notation, the model can be written as:

$$\mathbf{MA} = \mathbf{X}\boldsymbol{\beta}_k + \boldsymbol{\varepsilon} \qquad k \in \{1,\ldots,K\} \tag{3.8}$$

- **MA** is an $(N \times 1)$ vector of the math anxiety per student
- $\boldsymbol{\beta}_k$ is an (8×1) vector of coefficients for the explanatory variables, they differ for each k
- **X** is an $(N \times 8)$ matrix containing the explanatory variables
- $\boldsymbol{\varepsilon}$ is an $(N \times 1)$ vector of errors. We assume $\boldsymbol{\varepsilon}$ is normally distributed with normal distribution $\boldsymbol{\varepsilon} \sim N(0,\sigma^2\mathbf{I})$

Student-Teacher model

For the last model, we will look into the effect of the interpersonal relationship of students who have the same teacher. This is a two-level mixed-effects model.

$$\begin{aligned} MA_i = {} & \beta_{0j} + \beta_{1j}X_{1i} + \beta_{2j}X_{2i} + \beta_{3j}X_{3i} + \beta_{4j}X_{4i} + \beta_{5j}X_{5i} + \beta_{6j}X_{6i} + \beta_{7j}X_{7i} + \beta_{8j}X_{8i} + \varepsilon_{ij} \\ & i \in \{1,\ldots,N\}, j \in \{1,\ldots,J\} \end{aligned} \tag{3.9}$$

- We assume ε_{ij} has a normal distribution $\varepsilon_{ij} \sim N(0,\sigma^2)$
- Definitions of all variables are provided in Table 3.1

In matrix notation, the model can be written as:

$$\mathbf{MA} = \mathbf{X}\boldsymbol{\beta}_j + \boldsymbol{\varepsilon} \qquad j \in \{1,\ldots,J\} \tag{3.10}$$

- **MA** is an $(N \times 1)$ vector of the math anxiety per student
- $\boldsymbol{\beta}_j$ is an (8×1) vector of coefficients for the explanatory variables, they differ for each j
- **X** is an $(N \times 8)$ matrix containing the explanatory variables
- $\boldsymbol{\varepsilon}$ is an $(N \times 1)$ vector of errors. We assume $\boldsymbol{\varepsilon}$ is normally distributed with normal distribution $\boldsymbol{\varepsilon} \sim N(0,\sigma^2\mathbf{I})$

4

Results

After the data had been collected and the appropriate models had been selected, the models were fitted to the data. Section 4.1 presents the results for the Teacher model. Section 4.2 presents the results for the Class model. Section 4.3 presents the results for the Student models. Finally, Section 4.4 compares the models. The R code for the different models is attached in Appendix C

4.1. Teacher model

Before fitting the Teacher model, the model assumptions were assessed. The Teacher model was based on teacher-level averages, resulting in only six observations. Consequently, the results should be interpreted with caution, and assumption checks should be regarded as exploratory rather than definitive. The normality of the residuals was assessed using qqplots. The normality of the residuals is necessary to get an estimation of the slope coefficients. In addition, the distributions of the explanatory variables were inspected for potential outliers and extreme skewness. The check for normality of the explanatory variables can be found in Figure 4.1.
Figure 4.1 shows that the general enjoyment, the procedure-focused instruction and strict do not look like a normal distribution. Given the limited sample size ($N = 6$), no definitive conclusions regarding departures from normality could be drawn, and the analyses proceeded under the normality assumption.

The distribution of the residuals was also assessed for normality. In Figure 4.2 the normal qqplots of the response variable, math anxiety, can be found.
The qqplot contains only six observations, making it difficult to reliably assess normality, but this can only be ruled out if there are more data points. Therefore, normality of the response variable was assumed for the purposes of this analysis.

Another assumption concerns multicollinearity among the explanatory variables. In Table 4.1 the correlation coefficients of the explanatory variables can be found.
Table 4.1 shows strong correlations among the interpersonal-behavior variables ($X_1, ..., X_8$). Therefore, no model included more than one interpersonal-behavior variable. Between the other variables there is also correlation. Correlations exceeding an absolute value of 0.55 were considered indicative of problematic multicollinearity. This relatively conservative threshold was chosen because the Teacher model was fitted to only six observations, making the regression estimates particularly sensitive to correlations among predictors. In such cases, the corresponding variables were not included simultaneously in the same model.

To assess the relevance of the explanatory variables, each predictor was first analyzed in a separate simple linear regression model. In Table 4.2 the estimators and p-values of all the separate models are found. Table 4.2 shows that growth mindset was the only individually significant predictor. Several other interpersonal-behavior variables showed moderate evidence of association with math anxiety, although none reached conventional significance levels. Since few predictors were individually significant, we chose to create models with the explanatory variables with a p-value less than 0.50, which are anger, growth mindset and the variables from the student survey.

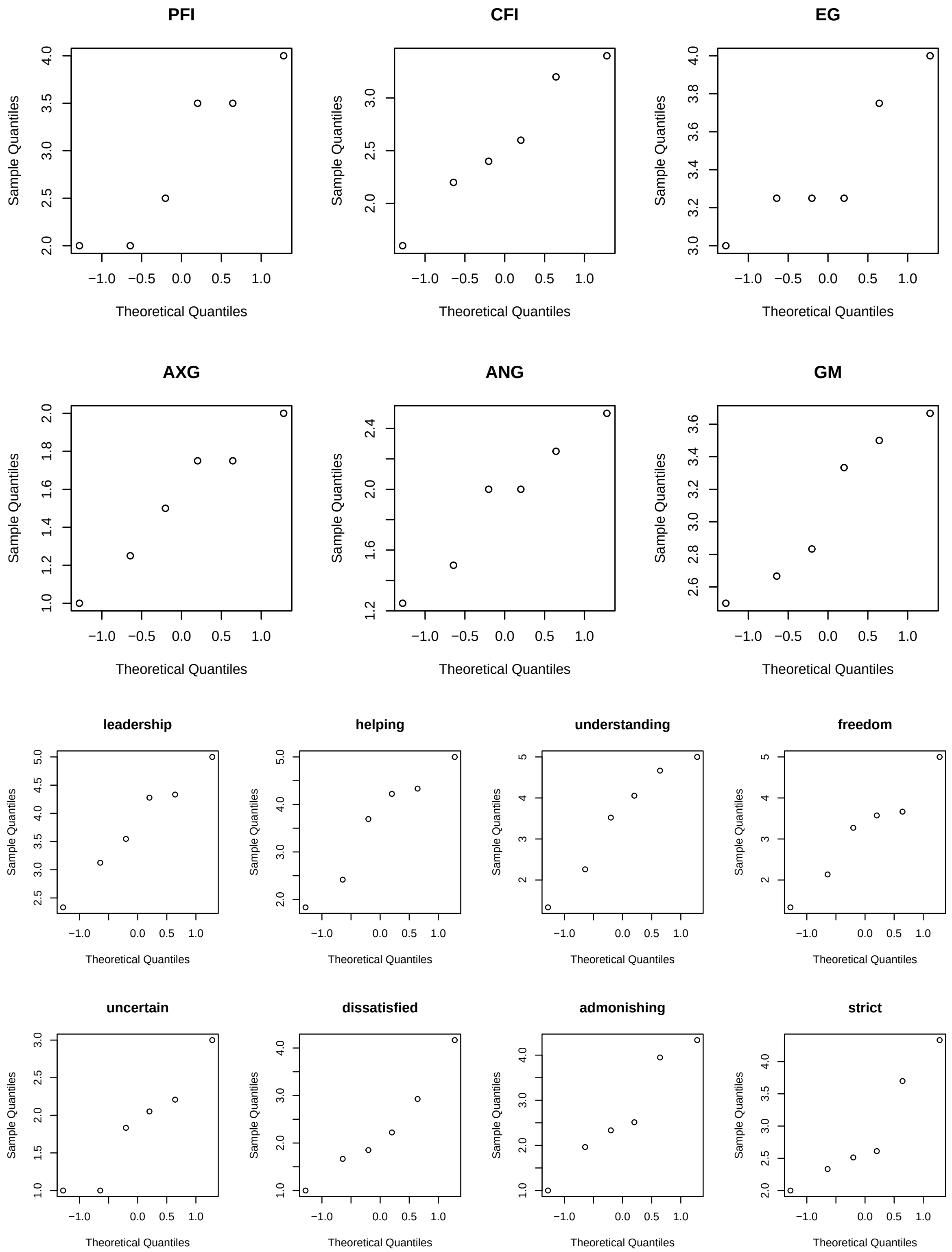

Figure 4.1: QQnorm plots for the explanatory variables of the Teacher model

	CFI	EG	AXG	ANG	GM	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
PFI	0.59	-0.10	-0.78	0.06	0.06	-0.49	-0,62	-0.65	-0.59	0.60	0.56	0.57	0.76
CFI		-0.17	-0.90	0.09	-0.30	-0.56	-0.52	-0.62	-0.53	0.80	0.62	0.40	0.66
EG			-0.15	-0.69	-0.28	0.56	0.56	0.59	0.42	-0.50	-0.54	-0.48	-0.61
ANX				0.24	0.26	0.36	0.39	0.47	0.42	-0.62	-0.44	-0.30	-0.55
ANG					-0.22	-0.80	-0.72	-0.68	-0.71	0.60	0.74	0.75	0.54
GM						0.26	0.17	0.16	0.31	-0.23	-0.35	-0.22	-0.12
X_1							0.98	0.98	0.98	-0.94	-0.98	-0.96	-0.88
X_2								0.99	0.98	-0.91	-0.98	-0.98	-0.94
X_3									0.97	-0.95	-0.98	-0.95	-0.96
X_4										-0.90	-0.98	-0.99	-0.88
X_5											0.94	0.84	0.90
X_6												0.95	0.93
X_7													0.87

Table 4.1: Correlation coefficients between the explanatory variables of the teacher model

	Estimate	p-value
PFI	0.09456	0.823
CFI	0.2826	0.600
EG	-0.2300	0.811
AXG	-0.2800	0.776
ANG	0.9803	0.149
GM	-1.1844	0.04085
Leadership	-0.5856	0.1036
Helping	-0.4205	0.1529
Understanding	-0.3364	0.1894
Freedom	-0.4619	0.07564
Uncertain	0.5943	0.220
Dissatisfied	0.5261	0.082
Admonishing	0.4523	0.1
Strict	0.4340	0.307

Table 4.2: Estimates and p-values for the separate explanatory variables for the Teacher model

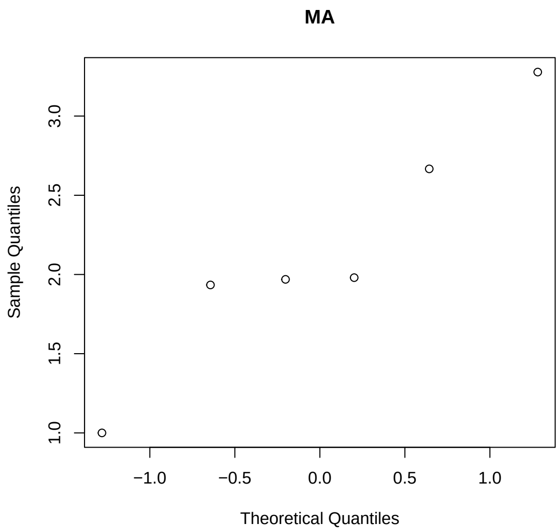

Figure 4.2: QQnorm plot for the response variable of the Teacher model

Twenty-one models were constructed. All of them based on the model in Equation 3.1. In Table 4.3 all the different models can be found, including the p-values of their explanatory variables, their normal and adjusted R^2, AIC and BIC. The p-values are distinguished into five levels of significance: "***" (p-value smaller than 0.001), "**" (p-value smaller than 0.01), "*" (p-value smaller than 0.05), "." (p-value smaller than 0.1), " " (p-value not significant). The results are ordered on increasing AIC.

Table 4.3 shows that the model with the smallest AIC, which has also the smallest BIC and the biggest adjusted R^2, contains the explanatory variables "growth mindset" and "admonishing". Another model that also performed well was the model containing "growth mindset" and "freedom", "leadership" or "helping". All those models have a relatively small AIC and BIC, and all achieved an R^2 close to 0.90, and have significant p-values for the explanatory variables.

Overall, the results suggest that the Teacher model may be useful for explaining variation in students' math anxiety. However, these findings should be interpreted with considerable caution because the model was fitted using only six teacher-level observations. Future research with a substantially larger dataset is required to validate the results, obtain more reliable parameter estimates, and increase statistical power.

4.2. Class model

Before fitting the Class model, the model assumptions were assessed. The distributions of the explanatory variables were inspected for severe departures from normality, while the normality assumption was assessed primarily through the residuals. The check for normality of the explanatory variables can be found in Figure 4.3.
Figure 4.3 shows that, the QQplots do not show substantial deviations from normality. So , for the rest of the model we use that the assumption is met.

Figure 4.4 presents a QQplot of the math anxiety scores. Although the observations appear to lie approximately on a straight line, the sample size is too small to draw definitive conclusions regarding normality. Therefore, no clear departures from normality were assumed for the purpose of this analysis.

Another assumption concerns multicollinearity among the explanatory variables. In Table 4.4 the correlation coefficients of the explanatory variables can be found.
Table 4.4 shows strong correlations between many of the explanatory variables. Correlations exceeding an absolute value of 0.55 were considered indicative of problematic multicollinearity. Given the small number of classes in the dataset, a relatively conservative threshold was adopted. Therefore, only combinations of

					p-value										
Expl. Var.	R^2	Adj. R^2	AIC	BIC	Int.	AXG	GM	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
GM, X_7	0.96	0.93	1.49	0.66	**		*							*	
ANX, GM, X_1	0.97	0.93	1.62	0.58	**		*	*							
GM, X_4	0.94	0.89	4.28	3.45	**		*				*				
GM, X_1	0.93	0.88	5.00	4.14	**		*	*							
GM, X_2	0.92	0.87	5.44	4.61	**		*		*						
GM, X_6	0.90	0.83	6.96	6.12	*		.						.		
GM, X_3	0.89	0.81	7.78	6.94	**		*			.					
GM, X_5	0.81	0.68	10.82	9.99	*		.								
GM, X_8	0.81	0.68	10.93	10.10	*		.								
GM	0.64	0.55	12.67	12.04	*		.								
X_4	0.59	0.48	13.51	12.88	**						.				
X_6	0.57	0.46	13.73	13.10									.		
X_7	0.53	0.41	14.26	13.64											
X_1	0.52	0.41	14.35	13.73	*										
ANX, GM	0.64	0.40	14.66	13.83	*										
X_2	0.44	0.30	15.37	14.74	*										
X_3	0.38	0.23	15.91	15.28	*										
ANX, X_1	0.53	0.22	16.26	15.43	.										
X_5	0.35	0.18	16.27	15.64											
X_8	0.25	0.07	17.05	16.43											
ANX	0.03	-0.21	18.61	17.99											

Table 4.3: R^2, adjusted R^2, AIC, BIC, and corresponding p-values of the different Teacher models

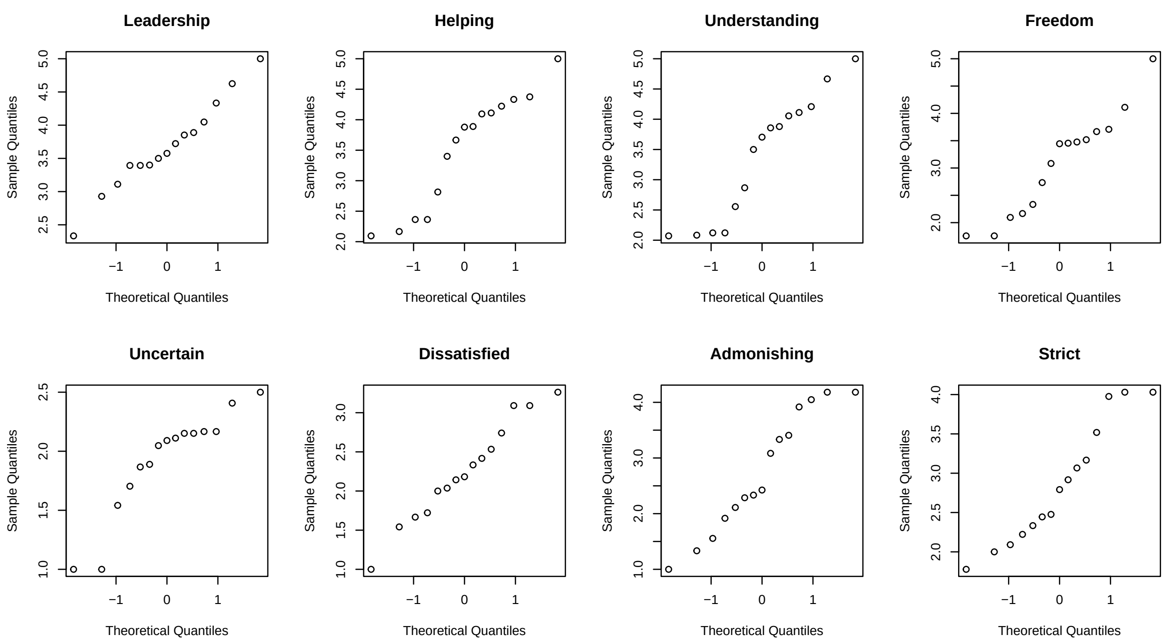

Figure 4.3: QQnorm plots for the explanatory variables of the Class model

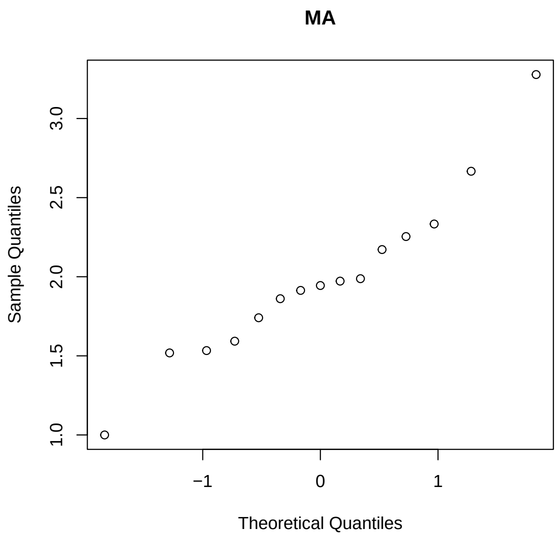

Figure 4.4: QQnorm plot for the math anxiety of the Class model

	Helping	Und.	Freedom	Uncertain	Diss.	Adm.	Strict
Lead.	0.87	0.85	0.79	-0.80	-0.78	-0.77	-0.49
Helping		0.97	0.94	-0.73	-0.91	-0.93	-0.80
Und.			0.95	-0.75	-0.93	-0.94	-0.83
Freedom				-0.71	-0.95	-0.95	-0.84
Uncertain					0.80	0.60	0.56
Diss.						0.91	0.83
Adm.							0.83

Table 4.4: Correlation coefficients between the explanatory variables of the Class model

variables with correlations below the chosen threshold were considered in the model selection process.

To assess the relevance of the explanatory variables, each predictor was first evaluated in a separate linear regression model. In Table 4.5 the estimators and p-values of all the separate models are found. None of the predictors reached the 5% significance level, although freedom and admonishing showed marginal significance ($p < 0.10$). Therefore, only freedom and admonishing were retained as candidate predictors for the subsequent model selection procedure. Four models were constructed. All four models were based on

	Estimate	p-value
Leadership	-0.2012	0.36948
Helping	-0.2119	0.18314
Understanding	-0.1660	0.264811
Freedom	-0.2591	0.0976
Uncertain	0.2472	0.4674
Dissatisfied	0.3700	0.1125
Admonishing	0.2267	0.09587
Strict	0.1760	0.3805

Table 4.5: Estimates and p-values for the separate explanatory variables for the Class model

Equation 3.3. In Table 4.6 all the different models can be found, including the p-values of their explanatory variable, their model and random R^2, AIC and BIC. The p-values are distinguished into five levels of significance: "***" (p-value smaller than 0.001), "**" (p-value smaller than 0.01), "*" (p-value smaller than 0.05), "." (p-value smaller than 0.1), " " (p-value not significant). The models are ordered by increasing AIC.

						p-value		
Fixed effects	Random components	R^2_m	R^2_c	AIC	BIC	Int.	X_4	X_7
X_7	Intercept only	0.21	0.84	26.20	29.03	**		*
X_4	Intercept only	0.14	0.78	28.12	30.95	***		
X_7	X_7	0.42	0.92	28.41	32.66	.		.
X_4	X_4	0.28	0.78	31.56	35.81	*		

Table 4.6: R^2_m, R^2_c, AIC, BIC and corresponding p-values of the different Class models

					p-value		
Explanatory variables	R^2	adj. R^2	AIC	BIC	Int.	X_4	X_7
X_4	0.20	0.14	25.38	27.50	**		.
X_7	0.20	0.14	25.41	27.54	***	.	

Table 4.7: R^2, adjusted R^2, AIC, BIC, and corresponding p-values of the different simplified Class models

The marginal R^2 values indicate that the fixed effects explain only a limited proportion of the variance (14-21%). In contrast, the conditional R^2 values show that including teacher-level random effects increases the explained variance to approximately 78-84%. This suggests that variation between teachers is considerably more important than variation explained by the class-level predictors. Because the two models with the smallest AIC only contains a random component in the intercept, we want to look into simple linear regression models for the Class-model too. The results of this can be found in Table 4.7.

Table 4.7 shows that neither simple regression model yielded significant predictor effects. In addition, both models had low R^2 values, indicating limited explanatory power.

To conclude, none of the class-level models provided strong evidence for significant fixed effects. A likely explanation is the limited number of classes per teacher. Because only two or three classes were observed per teacher, relatively little variation is available at the class level, whereas a larger proportion of the variance is captured by the teacher-level random effects. We want a good explanation for both, to do this we need to make sure there are more classes per teacher. This limitation is inherent to the available dataset.
The class-level fixed effects explain only a small proportion of the variation in math anxiety. In contrast, the large conditional R^2 values indicate that teacher-level differences account for a substantial proportion of the variance in math anxiety. These results suggest that a two-level mixed-effects model remains appropriate, but that substantially more classes per teacher would be required to reliably estimate class-level effects.

4.3. Student model

The next three models investigate the relationship between students' perceptions of teacher interpersonal behavior and math anxiety. The distributions of the explanatory variables were inspected for potential outliers and severe departures from normality. Figure 4.5 presents QQplots and boxplots of the explanatory variables. As we can see in the QQnorm plots is that there are a lot of points at the top and bottom of the plots. Several observations appear as potential outliers in the boxplots for leadership, uncertain, and dissatisfied. These observations likely reflect respondents' selecting extreme response categories, which is common in Likert-scale survey data. The reason for those points at the top and bottom is that some people tend to go for the extreme answers if they are biased (Manstein et al., 2023). Some students may hold particularly positive or negative perceptions of their teacher and are therefore more likely to go for the extreme answers that correspond with their opinion about their teacher. That explains the clustering of values at the top and bottom. Apart from a limited number of potential outliers, the distributions did not exhibit severe irregularities. Given the limited number of potential outliers and the overall shape of the distributions, these observations were not treated separately in the subsequent analyses.

The distribution of the residuals was inspected because severe skewness may affect the residual distribution of the regression model. Figure 4.6 presents qqplots of math anxiety and its logarithmic transformation. The qqplot of math anxiety shows some deviation from the theoretical normal distribution, particularly in the upper tail, suggesting that the variable is not perfectly normally distributed. A logarithmic transformation of math anxiety was therefore considered to reduce skewness. The qqplot of the transformed variable showed a closer approximation to normality. Consequently, the natural logarithm of math anxiety was used as the response variable in the Student model.

The last assumption is the independence of the multiple explanatory variables. In Table 4.8, the correlation coefficients of the variables are found. Table 4.8 shows strong correlations among several interpersonal-behavior dimensions, indicating the presence of substantial multicollinearity. Consequently, including multiple highly correlated predictors in the same model could lead to multicollinearity issues. Correlations exceeding an absolute value of 0.55 were considered indicative of problematic multicollinearity. Therefore, highly correlated explanatory variables were not included simultaneously in the same model.

	Helping	Und.	Freedom	Uncertain	Diss.	Adm.	Strict
Lead.	0.75	0.72	0.52	-0.69	-0.58	-0.55	-0.18
Helping		0.92	0.76	-0.51	-0.70	-0.73	-0.46
Und.			0.84	-0.50	-0.70	-0.75	-0.52
Freedom				-0.31	-0.59	-0.71	-0.61
Uncertain					0.62	0.41	0.24
Diss.						0.69	0.54
Adm.							0.71

Table 4.8: Correlation coefficients between the explanatory variables of the Student model

To assess the relevance of the explanatory variables, each predictor was first analyzed in a separate simple linear regression model. In Table 4.9 the estimators and p-values of all the separate models are found. In contrast to the Teacher and Class models, nearly all interpersonal-behavior dimensions were individually associated with math anxiety. Therefore, strict was excluded from the subsequent model-building procedure.

	Estimate	p-value
Leadership	-0.15784	$9.33 \cdot 10^{-5}$
Helping	-0.0.09931	0.00163
Understanding	-0.09923	0.0019
Freedom	-0.07898	0.0276
Uncertain	0.25107	$1 \cdot 10^{-7}$
Dissatisfied	0.14623	0.000138
Admonishing	0.07134	0.0277
Strict	$-6.622 \cdot 10^{-5}$	0.999

Table 4.9: Estimates and p-values for the separate explanatory variables for the student model

4.3.1. Three-level model

For the three-level model, 38 candidate models were constructed using explanatory variables that were individually significant and did not exhibit problematic multicollinearity. For each model, both random-intercept specifications including random slopes were considered. Table 4.10 summarizes the candidate models, including their marginal and conditional R^2 values, AIC, BIC, and predictor significance levels. The p-values are distinguished into five levels of significance: ***" (p-value smaller than 0.001), "**" (p-value smaller than 0.01), "*" (p-value smaller than 0.05), "." (p-value smaller than 0.1), " " (p-value not significant). The results are ordered by increasing AIC.

Table 4.10 shows that models with the lowest AIC and BIC values generally have relatively low marginal R^2 values and few significant fixed effects. This suggests that increasing model complexity does not necessarily improve explanatory power. The model containing leadership as the only fixed effect and a random inter-

						p-value							
Fix. eff.	Rand. comp.	R_m^2	R_c^2	AIC	BIC	Int.	X_1	X_4	X_7	X_2	X_5	X_3	X_6
X_5	Int.	0.26	0.29	59.24	71.96						***		
X_4, X_5	Int.	0.28	0.37	64.04	79.30	*					***		
X_3, X_5	Int.	0.28	0.37	64.27	79.53	.					***		
X_2, X_5	Int.	0.27	0.36	64.46	79.72	.					***		
X_5, X_7	Int.	0.26	0.32	65.79	81.05						***		
X_5, X_7	X_7	0.26	0.52	66.00	91.43						***		
X_5	X_5	0.25	0.30	66.82	89.71						***		
X_4, X_5	X_4	0.20	0.38	70.29	95.73						***		
X_1	Int.	0.16	0.26	71.28	84.00	***	***						
X_3, X_5	X_3	0.26	0.36	71.56	96.99	.					***		
X_4, X_5	X_5	0.28	0.37	71.82	97.25	*					***		
X_6	Int.	0.16	0.24	71.88	84.60	*							***
X_2, X_5	X_2	0.26	0.36	71.91	97.34	.					***		
X_3, X_5	X_5	0.27	0.37	72.08	97.51	.					***		
X_2, X_5	X_5	0.27	0.36	72.35	97.79	.					***		
X_3	Int.	0.18	0.36	73.33	86.04	***						***	
X_5, X_7	X_5	0.26	0.32	73.54	98.98						***		
X_2	Int.	0.17	0.33	74.08	86.80	***				***			
X_7	X_7	0.17	0.52	74.19	97.08	.			*				
X_1, X_7	X_7	0.21	0.55	74.90	100.34	**	*		.				
X_1, X_4	Int.	0.18	0.31	76.71	91.97	***	**						
X_1, X_7	Int.	0.18	0.31	77.07	92.33	***	**						
X_5, X_7	X_5, X_7	0.26	0.53	77.98	118.68				.		***		
X_6	X_6	0.15	0.27	78.34	101.23	*							**
X_2	X_2	0.15	0.40	78.69	101.58	***				*			
X_4	X_4	0.02	0.39	79.04	101.93	*							
X_1	X_1	0.16	0.26	79.16	102.05	***	***						
X_3	X_3	0.16	0.36	79.80	102.69	***						**	
X_4	Int.	0.09	0.23	79.94	92.66	***		**					
X_7	Int.	0.10	0.25	80.18	92.89	*			*				
X_1, X_4	X_4	0.09	0.37	81.43	106.87	***	*						
X_4, X_5	X_4, X_5	0.21	0.38	82.25	122.94						***		
X_3, X_5	X_3, X_5	0.26	0.36	83.52	124.21	.					***		
X_2, X_5	X_2, X_5	0.26	0.36	83.91	124.60	.					***		
X_1, X_4	X_1	0.19	0.32	84.51	109.95	***	**						
X_1, X_7	X_1, X_7	0.20	0.62	84.61	125.30	.							
X_1, X_7	X_1	0.18	0.31	85.03	110.47	***	**						
X_1, X_4	X_1, X_4	0.12	0.36	90.27	130.96	***							

Table 4.10: R_m^2, R_c^2, AIC, BIC, and corresponding p-values of the different three-level Student models

cept showed significant predictor effects but relatively low marginal R^2, indicating limited explanatory power at the fixed-effects level. Among the candidate models, the model including leadership and admonishing achieved a favorable balance between model fit (AIC/BIC), explained variance, and predictor significance, making it the most promising model.

A likely explanation for the limited significance of many models is the hierarchical structure of the dataset. Although 93 students participated, these students were distributed across only fourteen classes and six teachers. Consequently, the amount of information available at the class and teacher levels was limited, reducing the statistical power of the three-level analyses. To let this be a better model a larger number of students per class and more classes per teacher would improve the reliability of the model estimates.

4.3.2. Student-Class model

For the Student-Class model, 38 candidate models were constructed, using explanatory variables that were individually significant and did not exhibit problematic multicollinearity. Both random-intercept models and models including random slopes were considered. Table 4.11 summarizes the results of all candidate models. The p-values are classified into five levels of significance: ***" (p-value smaller than 0.001), "**" (p-value smaller than 0.01), "*" (p-value smaller than 0.05), "." (p-value smaller than 0.1), " " (p-value not significant). The results are ordered from lowest to highest AIC.

Table 4.11 shows that the models with the lowest AIC and BIC values generally have relatively low marginal R^2 values and only a limited number of significant fixed effects. The relatively small differences in AIC and BIC among the best-performing models suggest that adding additional explanatory variables yield only limited improvements in model fit. Although the leadership-only model contains significant predictor effects, its marginal R^2 remains relatively low. In contrast, the best-performing models according to AIC and BIC generally include admonishing behavior.
The marginal R^2 values ranged from approximately 0.02 to 0.28, indicating that the fixed effects explained a modest proportion of the variance. The conditional R^2 values were generally higher, suggesting that the random effects contributed additional explanatory power.

A likely explanation for the limited performance of the Student-Class models is the restricted amount of information available at the class level. In addition, classes taught by the same teacher are likely to be correlated, which violates the independence assumption underlying ordinary regression models. These results suggest a teacher-level variation should be included in the multilevel model because it may account for a substantial proportion of the variability in math anxiety.

4.3.3. Student-Teacher model

For the Student-Teacher model, 38 candidate models were constructed, using explanatory variables that were individually significant and did not exhibit problematic multicollinearity. Both random-intercept models and models including random slopes were considered. Table 4.12 summarizes the results of all candidate models. The p-values are distinguished into five levels of significance: ***" (p-value smaller than 0.001), "**" (p-value smaller than 0.01), "*" (p-value smaller than 0.05), "." (p-value smaller than 0.1), " " (p-value not significant). The results are ordered from lowest to highest AIC.

Table 4.12 shows that several models achieved relatively low AIC and BIC values while maintaining moderate marginal and conditional R^2 values. The higher conditional R^2 values indicate that incorporation teacher-level random effects explains additional variation in math anxiety beyond that explained by the fixed effects alone.

Overall, the Student-Teacher model performed well and showed moderate explanatory power. To further assess the contribution of the teacher level, corresponding single-level student models were also fitted. The results are presented in Table 4.13. Compared with the Student-Teacher models, the corresponding single-level models generally achieved lower R^2 values, indicating that the teacher-level random effects contributed additional explanatory power. The corresponding single-level models yielded weaker results, suggesting that incorporating a teacher level may improve the explanation of variation in math anxiety.

						p-value							
Fix. eff.	Rand. comp.	R^2_m	R^2_c	AIC	BIC	Int.	X_1	X_4	X_7	X_2	X_5	X_3	X_6
X_5	Int.	0.26	0.27	57.73	69.90						***		
X_5	X_5	0.26	0.27	61.72	76.98						***		
X_4, X_5	Int.	0.28	0.32	63.60	76.31						***		
X_3, X_5	Int.	0.28	0.31	63.79	76.50						***		
X_2, X_5	Int.	0.27	0.30	63.90	76.61						***		
X_5, X_7	Int.	0.26	0.28	64.74	77.46						***		
X_5, X_7	X_7	0.27	0.46	66.06	83.86						***		
X_4, X_5	X_4	0.25	0.32	67.31	85.11						***		
X_4, X_5	X_5	0.28	0.32	67.59	85.39						***		
X_3, X_5	X_3	0.27	0.31	67.77	85.58						***		
X_3, X_5	X_5	0.28	0.32	67.78	85.58						***		
X_2, X_5	X_5	0.27	0.31	67.87	85.68						***		
X_2, X_5	X_2	0.27	0.30	67.89	85.70						***		
X_5, X_7	X_5	0.26	0.28	68.94	86.54						***		
X_1	Int.	0.17	0.21	70.70	80.87	***	***						
X_5, X_7	X_5, X_7	0.28	0.47	71.41	96.84						***		
X_6	Int.	0.16	0.20	71.45	81.62	*							***
X_4, X_5	X_4, X_5	0.25	0.31	73.14	98.57						***		
X_3, X_5	X_3, X_5	0.27	0.31	73.75	99.18						***		
X_2, X_5	X_2, X_5	0.27	0.31	73.86	99.29						***		
X_1	X_1	0.16	0.21	74.63	89.89	***	***						
X_3	Int.	0.17	0.28	74.85	85.03	***						***	
X_1, X_7	X_7	0.22	0.53	75.35	93.16	***	**						
X_6	X_6	0.16	0.20	75.52	90.78	*							**
X_2	Int.	0.15	0.23	75.59	85.76	***				***			
X_1, X_4	Int.	0.18	0.26	76.69	89.40	***	**						
X_1, X_7	Int.	0.18	0.25	77.07	89.78	***	**						
X_1, X_7	X_1, X_7	0.20	0.57	77.50	102.93	.							
X_3	X_3	0.17	0.28	78.79	94.05	***						**	
X_2	X_2	0.14	0.28	78.96	94.22	**				*			
X_7	X_7	0.14	0.40	79.56	94.82	*			*				
X_1, X_4	X_4	0.13	0.27	79.91	97.71	***	**						
X_1, X_4	X_1	0.18	0.25	80.60	98.41	***	**						
X_1, X_7	X_1	0.17	0.25	80.89	98.70	***	**						
X_4	Int.	0.07	0.13	80.92	91.10	***		*					
X_4	X_4	0.02	0.27	80.99	96.25	*							
X_1, X_4	X_1, X_4	0.13	0.31	81.12	106.55	***							
X_7	Int.	0.07	0.12	81.40	91.57	**			*				

Table 4.11: R^2_m, R^2_c, AIC, BIC, and corresponding p-values of the different Student-Class models

						p-value							
Fix. eff.	Rand. comp.	R^2_m	R^2_c	AIC	BIC	Int.	X_1	X_4	X_7	X_2	X_5	X_3	X_6
X_5	Int.	0.27	0.28	57.65	67.82						***		
X_5	X_5	0.26	0.28	61.51	76.77						***		
X_3, X_5	Int.	0.29	0.42	61.60	74.31	*					***	.	
X_2, X_5	Int.	0.29	0.42	61.63	74.34	*				.	***		
X_4, X_5	Int.	0.29	0.38	62.22	74.93	*		.			***		
X_5, X_7	Int.	0.28	0.38	63.38	76.09						***		
X_5, X_7	X_7	0.33	0.49	64.29	82.10						***		
X_2, X_5	X_5	0.27	0.44	64.39	82.20	*				*	*		
X_3, X_5	X_5	0.27	0.44	64.63	82.44	*					*	*	
X_3, X_5	X_3	0.26	0.46	64.98	82.78	*					***		
X_4, X_5	X_5	0.27	0.39	65.34	83.15	.		.			*		
X_2, X_5	X_2	0.28	0.43	65.43	83.23	*					***		
X_4, X_5	X_4	0.27	0.41	65.72	83.52	.					***		
X_5, X_7	X_5, X_7	0.31	0.56	66.18	92.18								
X_5, X_7	X_5	0.26	0.40	66.60	84.40						*		
X_1	Int.	0.19	0.29	69.05	79.22	***	***						
X_2, X_5	X_2, X_5	0.27	0.44	70.36	95.80					.	.		
X_3, X_5	X_3, X_5	0.26	0.45	70.56	95.99	.					*	.	
X_4, X_5	X_4, X_5	0.26	0.44	71.02	96.45								
X_2	Int.	0.20	0.38	71.20	81.37	***				***			
X_3	Int.	0.20	0.39	71.52	81.69	***						***	
X_6	X_6	0.21	0.29	72.94	88.20	*							*
X_1	X_1	0.20	0.30	73.01	88.27	***	***						
X_1, X_7	X_7	0.30	0.54	73.20	91.01	***	**						
X_1, X_4	Int.	0.22	0.35	74.03	86.74	***	***						
X_2	X_2	0.23	0.39	74.88	90.14	**				**			
X_3	X_3	0.19	0.40	75.38	90.64	***						***	
X_7	X_7	0.26	0.49	76.55	91.81				.				
X_1, X_7	Int.	0.23	0.42	76.81	94.61	***	***						
X_1, X_7	X_1	0.23	0.41	77.01	94.81	***	***		.				
X_1, X_7	X_1, X_7	0.30	0.54	77.70	103.14								
X_1, X_4	X_4	0.22	0.35	78.02	95.83	***	***						
X_1, X_4	X_1	0.22	0.35	78.03	95.83	***	***						
X_7	Int.	0.14	0.30	78.44	88.61	.			**				
X_4	Int.	0.10	0.18	80.46	90.63	***		**					
X_1, X_4	X_1, X_4	0.23	0.38	83.88	109.31	***	.						
X_4	X_4	0.14	0.26	84.04	99.30	*							

Table 4.12: R^2_m, R^2_c, AIC, BIC, and corresponding p-values of the different Student-Teacher models

					p-value							
Exp. Var.	R^2	Adj. R^2	AIC	BIC	Int.	X_1	X_4	X_7	X_2	X5	X_3	X_6
X_5	0.27	0.26	46.18	53.81						***		
X_4, X_5	0.27	0.26	47.53	57.70						***		
X_2, X_5	0.27	0.25	47.62	57.79						***		
X_3, X_5	0.27	0.25	47.64	57.81						***		
X_5, X_7	0.27	0.25	48.14	58.32						***		
X_1	0.15	0.14	59.64	67.27	***	***						
X_6	0.15	0.14	60.40	68.03	**							***
X_1, X_4	0.15	0.14	61.54	71.72	***	**						
X_1, X_7	0.15	0.14	61.61	71.78	***	**						
X_2	0.10	0.09	65.13	72.76	***				**			
X_3	0.10	0.09	65.42	73.05	***						**	
X_4	0.05	0.04	70.34	77.97	***		*					
X_7	0.05	0.04	70.34	77.97	***			*				

Table 4.13: R^2, adjusted R^2, AIC, BIC, and corresponding p-values of the different simplified Student models

4.4. Comparison

The seven models differed substantially in explanatory power and model fit. The Teacher model achieved high adjusted R^2 values and identified several significant predictors. However, because the model was based on only six teacher-level observations, the results should be interpreted with caution. The Class model performed less well, with low marginal R^2 values and limited evidence for significant class-level effects.

The Student model showed strong associations between students' perceptions of teacher interpersonal behavior and math anxiety. Several interpersonal-behavior dimensions were individually significant predictors of students' math anxiety. The Student-Class model provided only moderate explanatory power, suggesting that class-level variation alone is insufficient to explain math anxiety. In contrast, the Student-Teacher model achieved moderate explanatory power and showed that teacher-level random effects contributed additional explained variance.

Overall, the Teacher model and the Student-Teacher model yielded the most promising results. The Teacher model suggests that teacher-level characteristics may be important predictors of math anxiety, whereas the Student-Teacher model combines teacher-level variation with a substantially larger sample size and a multi-level structure. Therefore, the Student-Teacher model and the Teacher model appears to be the most suitable models for future research on the influence of teachers on students' mathematics anxiety.

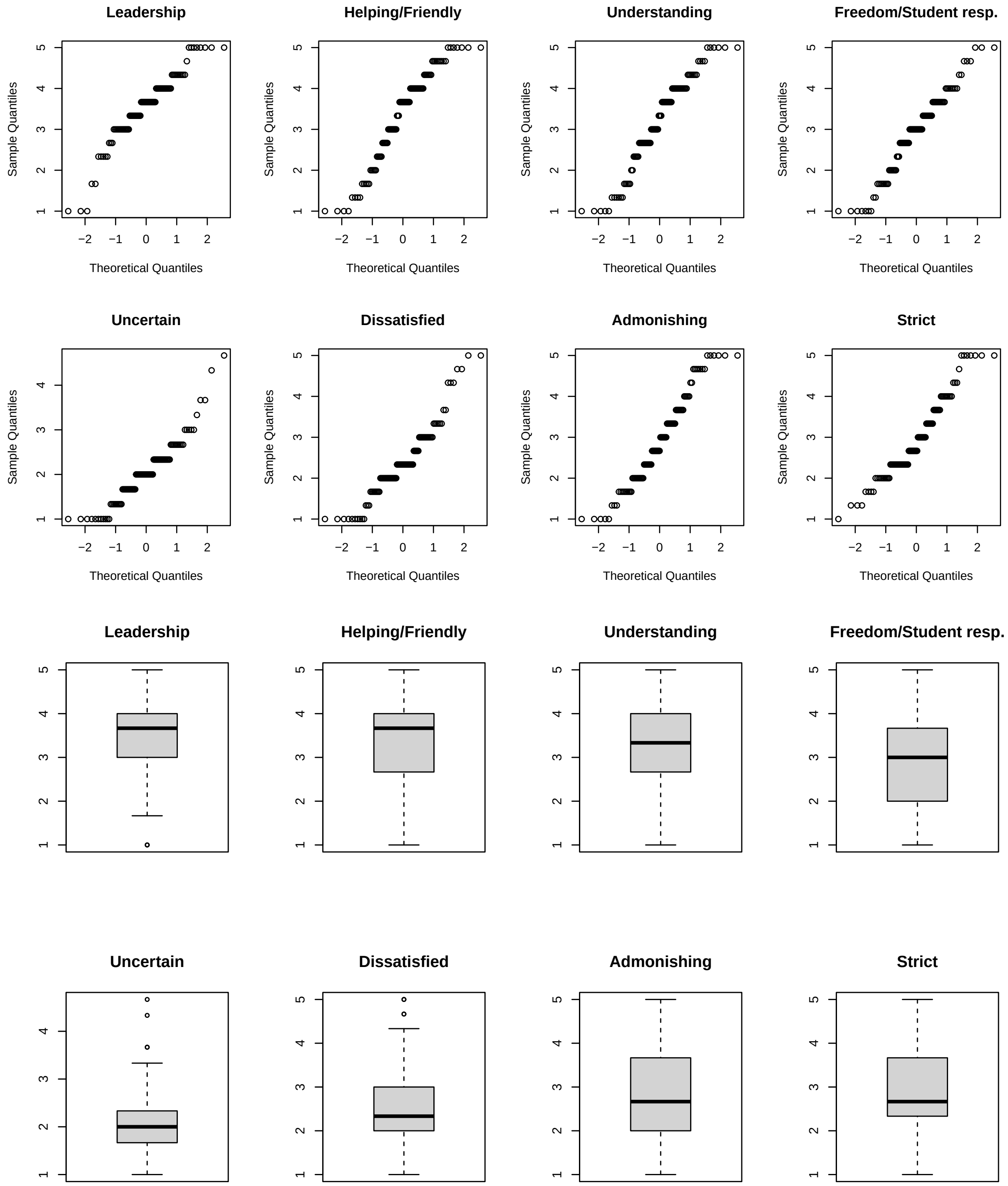

Figure 4.5: QQnorm and boxplots for the explanatory variables of the Student model

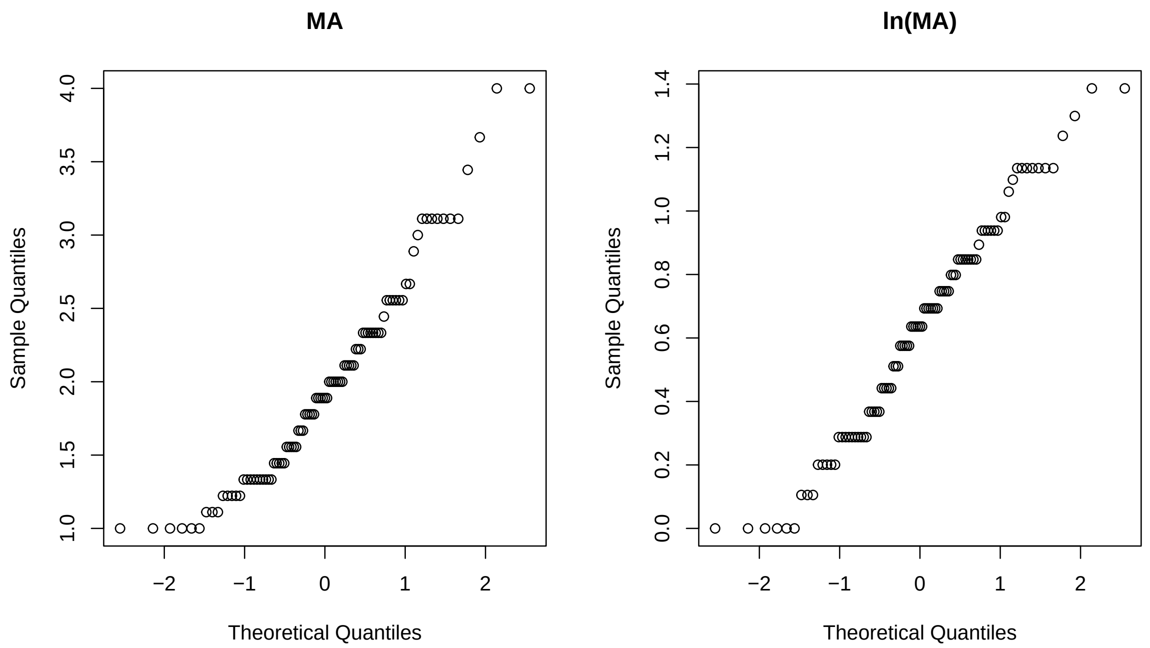

Figure 4.6: QQnorm plots for the math anxiety and natural logarithm of the math anxiety for the Student model

5
Conclusion/Discussion

The results suggest that the Teacher model and the Student-Teacher model are the most suitable models for analyzing the influence of math teachers on students' math anxiety. Of these two models, the Student-Teacher combines teacher-level effects with a substantially larger sample size and a multilevel structure. The Teacher model combines explanatory variables obtained by the student questionnaire as well as the teacher questionnaire.

The constructed Teacher model showed high explanatory power and provided the best fit among the teacher-level models. However, the model was based on only six teachers, which limits the reliability of the results. The Class model performed considerably worse, showing weak explanatory power and few significant effects. This suggests that class-level variation alone is insufficient tot explain differences in math anxiety.

The Student model explained a substantial proportion of the variation in math anxiety, indicating that student-level information is valuable for understanding math anxiety. However, this model does not account for the hierarchical structure of the data, where students are nested within classes and teachers.

The Student-Class model provided only limited additional explanatory value compared with the Student model. The relatively small number of students per class and dependence between classes taught by the same teacher likely reduced the effectiveness of this model. In contrast, the Student-Teacher model incorporated teacher-level variation and achieved moderate explanatory power while using substantially more observations than the Teacher model. The higher conditional R^2 values observed in this model suggest that teacher-level variation contributes meaningfully to explaining variation in mathematics anxiety.

Taken together, these results indicate that models that explicitly incorporate teacher-level variation are more suitable for studying math anxiety that models that focus solely on class-level variation. This finding supports the use of the Student-Teacher and Teacher models in future research. At the same time, the limited number of participating teachers means that the present results should be interpreted as exploratory rather than definitive.

Future research could consider generalized linear mixed-effects models, as the math anxiety scores did not fully satisfy the normality assumption. Such models may provide a more appropriate framework for analyzing math anxiety while accounting for the hierarchical structure of the data and the non-normal distribution of the response variable. In addition, larger datasets containing more teachers and more classes per teacher would allow more reliable estimation of teacher- and class-level effects.

Bibliography

Ashcraft, M. H. (2002). Math Anxiety: Personal, Educational, and Cognitive Consequences. *Current Directions in Psychological Science*, *11*(5), 181–185. https://doi.org/10.1111/1467-8721.00196

Beilock, S. L., & Maloney, E. A. (2015). Math Anxiety: A Factor in Math Achievement Not to Be Ignored. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 4–12. https://doi.org/10.1177/2372732215601438

Finch, W. H., Bolin, J. E., & Kelley, K. (2014). Multilevel Modeling Using R.

Fox, J. (2016). *Applied regression analysis and generalized linear models* (Third edition). SAGE Publications.

Frenzel, A. C., Daniels, L., & Burić, I. (2021). Teacher emotions in the classroom and their implications for students [_eprint: https://doi.org/10.1080/00461520.2021.1985501]. *Educational Psychologist*, *56*(4), 250–264. https://doi.org/10.1080/00461520.2021.1985501

Frenzel, A. C., Pekrun, R., Goetz, T., Daniels, L. M., Durksen, T. L., Becker-Kurz, B., & Klassen, R. M. (2016). Measuring Teachers' enjoyment, anger, and anxiety: The Teacher Emotions Scales (TES). *Contemporary Educational Psychology*, *46*, 148–163. https://doi.org/10.1016/j.cedpsych.2016.05.003

Greven, S., & Kneib, T. (2010). On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika*, *97*(4), 773–789. https://doi.org/10.1093/biomet/asq042

Klingberg, L. (2025). EU looks to plug STEM skills gap | Science|Business. Retrieved June 15, 2026, from https://sciencebusiness.net/news/careers/eu-looks-plug-stem-skills-gap

Laird, N. M., & Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics*, *38*(4), 963–974. https://doi.org/10.2307/2529876

Lapointe, J. M., Legault, F., & Batiste, S. J. (2005). Teacher interpersonal behavior and adolescents' motivation in mathematics: A comparison of learning disabled, average, and talented students. *International Journal of Educational Research*, *43*(1), 39–54. https://doi.org/10.1016/j.ijer.2006.03.005

Lee, J. (2025). Concept-Focused and Procedure-Focused Instruction on the Algebra Performance of Grade 9 Students With and Without Mathematics Difficulty. *Journal of Learning Disabilities*, *58*(3), 192–209. https://doi.org/10.1177/00222194241249960

Luttenberger, S., Wimmer, S., & Paechter, M. (2018). Spotlight on math anxiety. *Psychology Research and Behavior Management*, *11*, 311–322. https://doi.org/10.2147/PRBM.S141421

Lyu, J., Yu, X., Li, S., Li, R., Zhang, C., Wang, H., & Zhou, X. (2025). The Effect of Students' Perceived Cognitive Activation on Mathematics Performance Depends on the Grade, Performance Level, and Activation Intensity. *International Journal of Science and Mathematics Education*, *23*(7), 2455–2475. https://doi.org/10.1007/s10763-025-10559-6

Maestrini, L., Hui, F. K. C., & Welsh, A. H. (2025). Restricted Maximum Likelihood Estimation in Generalized Linear Mixed Models.

Manstein, S. M., Shiah, E., & Laikhter, E. (2023). Surveys and questionnaires. In *Translational Surgery* (pp. 273–280). Elsevier. https://doi.org/10.1016/B978-0-323-90300-4.00092-6

Matrix calculus [Page Version ID: 1315897806]. (2025, October). Retrieved June 1, 2026, from https://en.wikipedia.org/w/index.php?title=Matrix_calculus&oldid=1315897806

Radišić, J., Videnović, M., & Baucal, A. (2015). Math anxiety—contributing school and individual level factors. *European Journal of Psychology of Education*, *30*(1), 1–20. https://doi.org/10.1007/s10212-014-0224-7

Rights, J. D., & Sterba, S. K. (2023). R-squared Measures for Multilevel Models with Three or More Levels. *Multivariate Behavioral Research*, *58*(2), 340–367. https://doi.org/10.1080/00273171.2021.1985948

Schmitz, E. A., Salemink, E., Wiers, R. W., & Jansen, B. R. J. (2022). Test of Measurement Invariance, and Evidence for Reliability and Validity of AMAS Scores in Dutch Secondary School and University Students. *Journal of Psychoeducational Assessment*, *40*(5), 663–677. https://doi.org/10.1177/07342829221086141

Seo, E., & Lee, Y.-k. (2021). Stereotype Threat in High School Classrooms: How It Links to Teacher Mindset Climate, Mathematics Anxiety, and Achievement. *Journal of Youth and Adolescence*, *50*(7), 1410–1423. https://doi.org/10.1007/s10964-021-01435-x

Shen, N., & González, B. (2021, April). Bayesian Information Criterion for Linear Mixed-effects Models. Retrieved May 29, 2026, from https://arxiv.org/abs/2104.14725v1

Shoshani, A. (2021). Growth mindset in the maths classroom: A key to teachers' well-being and effectiveness. *Teachers and Teaching*, *27*(8), 730–752. https://doi.org/10.1080/13540602.2021.2007370

Snijders, T. A. B., & Bosker, R. (2011). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* [Open Library ID: OL28472018M]. SAGE Publications, Limited.

Sohil, F., Sohali, M. U., & Shabbir, J. (2022). An introduction to statistical learning with applications in R: By Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, New York, Springer Science and Business Media, 2013, $41.98, eISBN: 978-1-4614-7137-7. *Statistical Theory and Related Fields*, *6*(1), 87–87. https://doi.org/10.1080/24754269.2021.1980261

Wubbels, T., Brok, P. d., Claessens, L., Mainhard, T., & Tartwijk, J. v. (2024). *Klassenmanagement vanuit een interpersoonlijk perspectief* [OCLC: 1473931924]. Telos Uitgevers.

A

Questionnaires

A.1. Teacher Questionnaire

A.1.1. Instructional Objectives

English version
Think about the full duration of this academic year. How much emphasis are you placing on each of the following objectives? [no, minimal, moderate, heavy]

Dutch version
Denk na over het schooljaar tot nu toe. Hoeveel nadruk leg je op elk van de volgende doelstellingen? [Geen nadruk, Weinig nadruk, Redelijke nadruk, Veel nadruk]

	English version	Dutch version
IO1	Developing students' problem -solving skills	Het ontwikkelen van probleemoplossende vaardigheden van leerlingen.
IO2	Teaching students to reason mathematically	Leerlingen wiskundig laten redeneren
IO3	Teaching students how mathematics ideas connect with one another	Leerlingen inzicht geven in hoe wiskundige ideeën met elkaar samenhangen
IO4	Teaching students the logical structure of mathematics	Leerlingen inzicht geven in de logische structuur van de wiskunde
IO5	Teaching students to explain ideas in mathematics effectively	Leerlingen leren om ideeën in de wiskunde effectief uit te leggen
IO6	Developing students' computational skills	Het ontwikkelen van rekenvaardigheden van leerlingen
IO7	Teaching students to perform computations with speed and accuracy	Leerlingen leren om berekeningen snel en nauwkeurig uit te voeren

Table A.1: English and Dutch items of the instructional objectives questionnaire

A.1.2. Teacher Emotions

English version
Below you find a list of statements describing your experiences as a teacher. Please indicate your personal response to each of these statements by clicking the button that best represents your answer. [Strongly disagree, disagree, agree, strongly agree]
When answering the items below, please think of the first class that you teach every week (for example the first class on Monday).

Dutch version
Hieronder staan een aantal uitspraken over jouw ervaringen als docent. Geef bij elke uitspraak aan in hoeverre deze op jou van toepassing is door het antwoord aan te kruisen dat jouw persoonijke beleving het beste

weergeeft.
Denk aan de eerst klas die je deze week lesgeeft (bijvoorbeeld je eerste les op maandag). Beantwoord de vragen met deze klas in gedachten.

	English version	Dutch version
EG1	I generally enjoy teaching	Over het algemeen vind ik lesgeven leuk
EG2	I generally have so much fun teaching that I gladly prepare and teach my lessons	Over het algemeen vind ik lesgeven zo leuk dat ik mijn lessen graag voorbereid en geef.
EG3	I often have reasons to be happy while I teach	Tijdens het lesgeven heb ik vaak redenen om blij te zijn.
EG4	I generally teach with enthusiasm	Over het algemeen geef ik les met enthousiasme
AXG1	I generally feel tense and nervous while teaching	Tijdens het lesgeven ben ik over het algemeen gespannen en zenuwachtig
AXG2	I am often worried that my teaching isn't going so well	Ik maak me vaak zorgen dat het lesgeven niet goed verloopt.
AXG3	Preparing to teach often causes me to worry	De voorbereiding van mijn lessen bezorgt mij vaak zorgen.
AXG4	I feel uneasy when I think about teaching	Als ik aan lesgeven denk, voel ik me onrustig
AGG1	I often have reasons to be angry while I teach	Tijdens het lesgeven heb ik vaak redenen om boos te zijn
AGG2	I often feel annoyed while teaching	Ik voel me vaak geïrriteerd tijdens het lesgeven
AGG3	Sometimes I get really mad while I teach	Soms wordt ik echt boos tijdens het lesgeven
AGG4	Teaching generally frustrates me	Lesgeven frustreert mij over het algemeen
EC1	I enjoy teaching these students	In deze klas vind ik lesgeven leuk
EC2	I have so much fun teaching these students that I gladly prepare and teach my lessons	In deze klas vind ik lesgeven zo leuk dat ik mijn lessen graag voorbereid en geef.
EC3	I teach these students with enthusiasm	In deze klas geef ik les met enthousiasme
EC4	I often have reason to be happy while I teach these students	Tijdens het lesgeven in deze klas heb ik vaak redenen om blij te zijn
AXC1	I feel tense and nervous while teaching these students	Tijdens het lesgeven in deze klas ben ik gespannen en zenuwachtig
AXC2	I am often worried that my teaching isn't going so well with these students	Ik maak me vaak zorgen dat het lesgeven in deze klas niet goed verloopt
AXC3	Preparing to teach these students often causes me to worry	De voorbereiding van lessen voor deze klas bezorgt mij vaak zorgen.
AXC4	I feel uneasy when I think about teaching these students	Als ik denk aan het lesgeven in deze klas, voel ik me onrustig
AGC1	I often have reason to be angry while I teach these students	Tijdens het lesgeven in deze klas heb ik vaak redenen om boos te zijn
AGC2	I often feel annoyed while teaching these students	Ik voel me vaak geïrriteerd tijdens het lesgeven in deze klas.
AGC3	Sometimes I get really mad at these students	In deze klas word ik soms echt boos tijdens het lesgeven
AGC4	Teaching these students frustrates me	In deze klas frustreert het lesgeven mij

Table A.2: English and Dutch items of the teacher emotions questionnaire

A.1.3. Growth Mindset

English version

Below are a number of statements about your mindset as a teacher in general. For each statement, indicate to what extent it applies to you by checking the answer that best reflects your personal experience. [Strongly disagree, disagree, disagree a little, agree a little, agree, strongly agree]

Dutch version

Hieronder staan een aantal uitspraken over jouw mindset als docent in het algemeen. Geef bij elke uitspraak aan in hoeverre deze op jouw van toepassing is door het antwoord aan te kruisen dat jouw persoonlijke beleving het beste weergeeft. [Helemaal mee oneens, oneens, beetje oneens, beetje eens, eens, helemaal eens]

	English version	Dutch version
GM1	You have a certain amount of intelligence, and you can't really do much to change it.	Je hebt een bepaalde hoeveelheid intelligentie, en je kunt er niet veel aan doen om dit te veranderen.
GM2	Your intelligence is something about you that you can't change very much.	Je intelligentie is iets van jezelf dat je niet veel kunt veranderen.
GM3	You can learn new things, but you can't really change your basic intelligence	Je kunt misschien wel nieuwe dingen leren, maar je kunt je basis intelligentie niet echt veranderen.
GM4	Your math ability is something about you that you can't change very much	Je wiskundige vaardigheid is iets van jezelf dat je niet veel kunt veranderen.
GM5	Most people can do well in math	De meeste mensen zijn goed in wiskunde.
GM6	High math abilities are innate	Sterke wiskundige vaardigheden zijn aangeboren.

Table A.3: English and Dutch items of the growth mindset questionnaire

A.2. Student questionnaire

A.2.1. Math anxiety

English version

Rate each item below in terms of how anxious you would feel during the event specified. [not anxious, a bit anxious, somewhat anxious, really anxious]

Dutch version

Geef bij iedere vraag aan hoe angstig je je zou voelen als die situatie echt gebeurt. [(bijna) niet angstig, een beetje angstig, middelmatig angstig, erg angstig]

	English version	Dutch version
AM1	Having to use the tables in the back of a mathematics book.	Je tabellen in een wiskundeboek moet aflezen
AM2	Thinking about an upcoming mathematics test one day before.	je een dag van tevoren denkt aan een opgegeven toets voor wiskunde.
AM3	Watching a teacher work an algebraic equation on the blackboard.	je kijkt naar een docent die een algebraïsche vergelijking uitwerkt op het bord.
AM4	Taking an examination in a mathematics course.	je een toets voor wiskunde moet maken
AM5	Being given a homework assignment of many difficult problems which is due the next class meeting.	je huiswerk met veel moeilijke opgaven krijgt die de volgende les af moeten zijn
AM6	Listening to a lecture in mathematics class.	je luistert naar uitleg in de wiskundeles
AM7	Listening to another student explain a mathematics formula.	je luistert naar andere leerlingen die wiskundige formules uitleggen
AM8	Being given a "pop" quiz in a mathematics class.	er een onverwachte SO is tijdens de wiskundeles
AM9	Starting a new chapter in a mathematics book.	je begint met een nieuw hoofdstuk in een wiskundeboek

Table A.4: English and Dutch items of the math anxiety questionnaire

A.2.2. Interpersonal relationship

English version

Below are 24 statements about your math teacher from whom you are taking lessons this school year. For each sentence, check which circle best matches your opinion of your math teacher.

Dutch version
Hieronder staan 24 uitspraken over jouw wiskundedocent van wie je dit schooljaar les krijgt. Kruis bij elke zin aan welk bolletje het beste past bij jouw mening over jouw wiskundedocent.

	English version	Dutch version
$X_1 1$	This teacher is a good leader	Deze leerkracht kan goed leidinggeven
$X_1 2$	This teacher has authority	Deze leerkracht heeft gezag
$X_1 3$	This teacher acts confidently	Deze leerkracht treedt zelfverzekerd op
$X_2 1$	This teacher is someone you can depend on.	Deze leerkracht is iemand op wie je kunt vertrouwen
$X_2 2$	This teacher has a sense of humor	Deze leerkracht heeft gevoel voor humor
$X_2 3$	This teacher's class is pleasant	Deze leerkracht heeft een prettige sfeer in de klas
$X_3 1$	This teacher is patient	Deze leerkracht is geduldig
$X_3 2$	This teacher empathizes with students	Deze leerkracht leeft mee met leerlingen
$X_3 3$	This teacher is flexible with students.	Deze leerkracht is soepel voor leerlingen
$X_4 1$	This teacher lets students go their own way.	Deze leerkracht laat leerlingen hun gang gaan
$X_4 2$	This teacher gives students what they want.	Deze leerkracht geeft leerlingen hun zin
$X_4 3$	This teacher allows a lot	Deze leerkracht vindt veel goed
$X_5 1$	This teacher is lax.	Deze leerkracht treedt slap op
$X_5 2$	This teacher seems uncertain.	Deze leerkracht maakt een onzekere indruk
$X_5 3$	This teacher is hesistant	Deze leerkracht treedt aarzelend op
$X_6 1$	This teacher is in a bad mood	Deze leerkracht is uit zijn/haar humeur
$X_6 2$	This teacher seems dissatisfied.	Deze leerkracht is ontevreden
$X_6 3$	This teacher makes a gloomy impression.	Deze leerkracht maakt een sombere indruk
$X_7 1$	This teacher can get angry	Deze leerkracht kan kwaad worden
$X_7 2$	This teacher threatens punishment	Deze leerkracht dreigt met straf
$X_7 3$	It is easy to pick a fight with this teacher	Deze leerkracht is driftig
$X_8 1$	This teacher decides whether students are allowed to say something.	Deze leerkracht bepaald of leerlingen iets mogen zeggen
$X_8 2$	We have to be silent in this teacher's class.	Bij deze leerkracht moet het stil zijn in de les
$X_8 3$	This teacher maintains strict order.	Deze leerkracht houdt streng orde

Table A.5: English and Dutch items of the interpersonal relationship questionnaire

B

Data Sample

Teacher	Class	Number of observations
T01	Class_01	9
	Class_02	4
	Class_03	9
	Class_04	6
	Class_05	11
T02	Class_06	3
	Class_07	8
	Class_08	7
T03	Class_09	4
	Class_10	5
	Class_11	14
T04	Class_13	1
T05	Class_15	1
T06	Class_17	11

Table B.1: Data sample in the pilot study

C

R codes

C.1. Teacher model

```r
library(dplyr)

#Create the mean values of the explanatory and response variables
MA_avt <- student_survey %>%
  group_by(Teacher_ID) %>%
    summarise(average = mean(MA, na.rm = TRUE))

leadership_avt <- student_survey %>%
  group_by(Teacher_ID) %>%
  summarise(average = mean(leadership, na.rm = TRUE))

helping_avt <- student_survey %>%
  group_by(Teacher_ID) %>%
  summarise(average = mean(helping, na.rm = TRUE))

understanding_avt <- student_survey %>%
  group_by(Teacher_ID) %>%
  summarise(average = mean(understanding, na.rm = TRUE))

freedom_avt <- student_survey %>%
  group_by(Teacher_ID) %>%
  summarise(average = mean(freedom, na.rm = TRUE))

uncertain_avt <- student_survey %>%
  group_by(Teacher_ID) %>%
  summarise(average = mean(uncertain, na.rm = TRUE))

dissatisfied_avt <- student_survey %>%
  group_by(Teacher_ID) %>%
  summarise(average = mean(dissatisfied, na.rm = TRUE))

admonishing_avt <- student_survey %>%
  group_by(Teacher_ID) %>%
  summarise(average = mean(admonishing, na.rm = TRUE))

strict_avt <- student_survey %>%
  group_by(Teacher_ID) %>%
  summarise(average = mean(strict, na.rm = TRUE))
```

```
PFI_avt <- teacher_survey %>%
  group_by(Teacher_ID) %>%
  summarise(average = mean(PFI, na.rm = TRUE))

CFI_avt <- teacher_survey %>%
  group_by(Teacher_ID) %>%
  summarise(average = mean(CFI, na.rm = TRUE))


EG_avt <- teacher_survey %>%
  group_by(Teacher_ID) %>%
  summarise(average = mean(EG, na.rm = TRUE))

AXG_avt <- teacher_survey %>%
  group_by(Teacher_ID) %>%
  summarise(average = mean(AXG, na.rm = TRUE))

ANG_avt <- teacher_survey %>%
  group_by(Teacher_ID) %>%
  summarise(average = mean(ANG, na.rm = TRUE))

GM_avt <- teacher_survey %>%
  group_by(Teacher_ID) %>%
  summarise(average = mean(GM, na.rm = TRUE))

#Check for normality of the response and explanatory variables
par(mfrow=c(1,1))
qqnorm(MA_avt$average, main = "MA")

par(mfrow = c(2,3))
qqnorm(PFI_avt$average, main = "PFI")
qqnorm(CFI_avt$average, main = "CFI")
qqnorm(EG_avt$average, main = "EG")
qqnorm(AXG_avt$average, main = "AXG")
qqnorm(ANG_avt$average, main = "ANG")
qqnorm(GM_avt$average, main = "GM")

par(mfrow = c(2,4))
qqnorm(leadership_avt$average, main = "leadership")
qqnorm(helping_avt$average, main = "helping")
qqnorm(understanding_avt$average, main = "understanding")
qqnorm(freedom_avt$average, main = "freedom")
qqnorm(uncertain_avt$average, main = "uncertain")
qqnorm(dissatisfied_avt$average, main = "dissatisfied")
qqnorm(admonishing_avt$average, main = "admonishing")
qqnorm(strict_avt$average, main = "strict")

#Check for correlation between the explanatory variables
cor(PFI_avt$average, CFI_avt$average)
cor(PFI_avt$average, EG_avt$average)
cor(PFI_avt$average, AXG_avt$average)
cor(PFI_avt$average, ANG_avt$average)
cor(PFI_avt$average, GM_avt$average)
cor(PFI_avt$average, leadership_avt$average)
cor(PFI_avt$average, helping_avt$average)
```

```
cor(PFI_avt$average, understanding_avt$average)
cor(PFI_avt$average, freedom_avt$average)
cor(PFI_avt$average, uncertain_avt$average)
cor(PFI_avt$average, dissatisfied_avt$average)
cor(PFI_avt$average, admonishing_avt$average)
cor(PFI_avt$average, strict_avt$average)

cor(CFI_avt$average, EG_avt$average)
cor(CFI_avt$average, AXG_avt$average)
cor(CFI_avt$average, ANG_avt$average)
cor(CFI_avt$average, GM_avt$average)
cor(CFI_avt$average, leadership_avt$average)
cor(CFI_avt$average, helping_avt$average)
cor(CFI_avt$average, understanding_avt$average)
cor(CFI_avt$average, freedom_avt$average)
cor(CFI_avt$average, uncertain_avt$average)
cor(CFI_avt$average, dissatisfied_avt$average)
cor(CFI_avt$average, admonishing_avt$average)
cor(CFI_avt$average, strict_avt$average)

cor(EG_avt$average, AXG_avt$average)
cor(EG_avt$average, ANG_avt$average)
cor(EG_avt$average, GM_avt$average)
cor(EG_avt$average, leadership_avt$average)
cor(EG_avt$average, helping_avt$average)
cor(EG_avt$average, understanding_avt$average)
cor(EG_avt$average, freedom_avt$average)
cor(EG_avt$average, uncertain_avt$average)
cor(EG_avt$average, dissatisfied_avt$average)
cor(EG_avt$average, admonishing_avt$average)
cor(EG_avt$average, strict_avt$average)

cor(AXG_avt$average, ANG_avt$average)
cor(AXG_avt$average, GM_avt$average)
cor(AXG_avt$average, leadership_avt$average)
cor(AXG_avt$average, helping_avt$average)
cor(AXG_avt$average, understanding_avt$average)
cor(AXG_avt$average, freedom_avt$average)
cor(AXG_avt$average, uncertain_avt$average)
cor(AXG_avt$average, dissatisfied_avt$average)
cor(AXG_avt$average, admonishing_avt$average)
cor(AXG_avt$average, strict_avt$average)

cor(ANG_avt$average, GM_avt$average)
cor(ANG_avt$average, leadership_avt$average)
cor(ANG_avt$average, helping_avt$average)
cor(ANG_avt$average, understanding_avt$average)
cor(ANG_avt$average, freedom_avt$average)
cor(ANG_avt$average, uncertain_avt$average)
cor(ANG_avt$average, dissatisfied_avt$average)
cor(ANG_avt$average, admonishing_avt$average)
cor(ANG_avt$average, strict_avt$average)

cor(GM_avt$average, leadership_avt$average)
cor(GM_avt$average, helping_avt$average)
cor(GM_avt$average, understanding_avt$average)
```

```r
cor(GM_avt$average, freedom_avt$average)
cor(GM_avt$average, uncertain_avt$average)
cor(GM_avt$average, dissatisfied_avt$average)
cor(GM_avt$average, admonishing_avt$average)
cor(GM_avt$average, strict_avt$average)

cor(leadership_avt$average, helping_avt$average)
cor(leadership_avt$average, understanding_avt$average)
cor(leadership_avt$average, freedom_avt$average)
cor(leadership_avt$average, uncertain_avt$average)
cor(leadership_avt$average, dissatisfied_avt$average)
cor(leadership_avt$average, admonishing_avt$average)
cor(leadership_avt$average, strict_avt$average)

cor(helping_avt$average, understanding_avt$average)
cor(helping_avt$average, freedom_avt$average)
cor(helping_avt$average, uncertain_avt$average)
cor(helping_avt$average, dissatisfied_avt$average)
cor(helping_avt$average, admonishing_avt$average)
cor(helping_avt$average, strict_avt$average)

cor(understanding_avt$average, freedom_avt$average)
cor(understanding_avt$average, uncertain_avt$average)
cor(understanding_avt$average, dissatisfied_avt$average)
cor(understanding_avt$average, admonishing_avt$average)
cor(understanding_avt$average, strict_avt$average)

cor(freedom_avt$average, uncertain_avt$average)
cor(freedom_avt$average, dissatisfied_avt$average)
cor(freedom_avt$average, admonishing_avt$average)
cor(freedom_avt$average, strict_avt$average)

cor(uncertain_avt$average, dissatisfied_avt$average)
cor(uncertain_avt$average, admonishing_avt$average)
cor(uncertain_avt$average, strict_avt$average)

cor(dissatisfied_avt$average, admonishing_avt$average)
cor(dissatisfied_avt$average, strict_avt$average)

cor(admonishing_avt$average, strict_avt$average)

#Create the model
Model_teacher <- lm(MA_avt$average ~ PFI_avt$average + CFI_avt$average
                    + AXG_avt$average + ANG_avt$average + GM_avt$average
                    + leadership_avt$average + helping_avt$average
                    + understanding_avt$average + freedom_avt$average
                    + uncertain_avt$average + dissatisfied_avt$average
                    + admonishing_avt$average + strict_avt$average)

#Print the model results
summary(Model_teacher)
AIC(Model_teacher)
BIC(Model_teacher)
```

C.2. Class Model

```
library(dplyr)
library(readxl)
library(Matrix)
library(lme4)
library(coda)
library(languageR)
library(lmerTest)
library(purrr)
library(MuMIn)

#Create the mean values for the response and explanatory variables
MA_avc <- student_survey %>%
  group_by(Class_ID, Teacher_ID) %>%
  summarise(average = mean(MA))

MA_avc <- MA_avc %>%
  rename(MA_gem = average)

leadership_avc <- student_survey %>%
  group_by(Class_ID) %>%
  summarise(average = mean(leadership))

leadership_avc <- leadership_avc %>%
  rename(leadership_gem = average)

helping_avc <- student_survey %>%
  group_by(Class_ID) %>%
  summarise(average = mean(helping))

helping_avc <- helping_avc %>%
  rename(helping_gem = average)

understanding_avc <- student_survey %>%
  group_by(Class_ID) %>%
  summarise(average = mean(understanding))

understanding_avc <- understanding_avc %>%
  rename(understanding_gem = average)

freedom_avc <- student_survey %>%
  group_by(Class_ID) %>%
  summarise(average = mean(freedom))

freedom_avc <- freedom_avc %>%
  rename(freedom_gem = average)

uncertain_avc <- student_survey %>%
  group_by(Class_ID) %>%
  summarise(average = mean(uncertain))

uncertain_avc <- uncertain_avc %>%
  rename(uncertain_gem = average)

dissatisfied_avc <- student_survey %>%
  group_by(Class_ID) %>%
  summarise(average = mean(dissatisfied))
```

```
dissatisfied_avc <- dissatisfied_avc %>%
  rename(dissatisfied_gem = average)

admonishing_avc <- student_survey %>%
  group_by(Class_ID) %>%
  summarise(average = mean(admonishing))

admonishing_avc <- admonishing_avc %>%
  rename(admonishing_gem = average)

strict_avc <- student_survey %>%
  group_by(Class_ID) %>%
  summarise(average = mean(strict))

strict_avc <- strict_avc %>%
  rename(strict_gem = average)

datasets <- list(MA_avc, leadership_avc, helping_avc, understanding_avc, freedom_avc,
                 uncertain_avc, dissatisfied_avc, admonishing_avc, strict_avc)

combined <- reduce(datasets, left_join, by = "Class_ID")

#Check for normality of the response and explanatory variables
par(mfrow=c(2,4))
qqnorm(combined$leadership_gem, main = "Leadership")
qqnorm(combined$helping_gem, main = "Helping")
qqnorm(combined$understanding_gem, main = "Understanding")
qqnorm(combined$freedom_gem, main = "Freedom")
qqnorm(combined$uncertain_gem, main = "Uncertain")
qqnorm(combined$dissatisfied_gem, main = "Dissatisfied")
qqnorm(combined$admonishing_gem, main = "Admonishing")
qqnorm(combined$strict_gem, main = "Strict")

par(mfrow=c(1,1))
qqnorm(combined$MA_gem, main = "MA")

#Check for correlation between the explanatory variables
cor(combined$leadership_gem, combined$helping_gem)
cor(combined$leadership_gem, combined$understanding_gem)
cor(combined$leadership_gem, combined$freedom_gem)
cor(combined$leadership_gem, combined$uncertain_gem)
cor(combined$leadership_gem, combined$dissatisfied_gem)
cor(combined$leadership_gem, combined$admonishing_gem)
cor(combined$leadership_gem, combined$strict_gem)

cor(combined$helping_gem, combined$understanding_gem)
cor(combined$helping_gem, combined$freedom_gem)
cor(combined$helping_gem, combined$uncertain_gem)
cor(combined$helping_gem, combined$dissatisfied_gem)
cor(combined$helping_gem, combined$admonishing_gem)
cor(combined$helping_gem, combined$strict_gem)

cor(combined$understanding_gem, combined$freedom_gem)
cor(combined$understanding_gem, combined$uncertain_gem)
cor(combined$understanding_gem, combined$dissatisfied_gem)
```

```
cor(combined$understanding_gem, combined$admonishing_gem)
cor(combined$understanding_gem, combined$strict_gem)

cor(combined$freedom_gem, combined$uncertain_gem)
cor(combined$freedom_gem, combined$dissatisfied_gem)
cor(combined$freedom_gem, combined$admonishing_gem)
cor(combined$freedom_gem, combined$strict_gem)

cor(combined$uncertain_gem, combined$dissatisfied_gem)
cor(combined$uncertain_gem, combined$admonishing_gem)
cor(combined$uncertain_gem, combined$strict_gem)

cor(combined$dissatisfied_gem, combined$admonishing_gem)
cor(combined$dissatisfied_gem, combined$strict_gem)

cor(combined$admonishing_gem, combined$strict_gem)

#Create the two-level mixed-effects model
Model_class <- lmer(MA_gem ~ leadership_gem + helping_gem + understanding_gem
                    + freedom_gem + uncertain_gem + dissatisfied_gem
                    + admonishing_gem + strict_gem + (leadership_gem
                                                      + helping_gem
                                                      + understanding_gem
                                                      + freedom_gem
                                                      + uncertain_gem
                                                      + dissatisfied_gem
                                                      + admonishing_gem
                                                      + strict_gem|Teacher_ID)
                    , data = combined)

#Print the model results
summary(Model_class)
r.squaredGLMM(Model_class)
AIC(Model_class)
BIC(Model_class)

#Create the standard linear model
Model_class_standard <- lm(MA_gem ~ leadership_gem + helping_gem
                           + understanding_gem + freedom_gem + uncertain_gem
                           + dissatisfied_gem + admonishing_gem + strict_gem
                           , data = combined)

#Print the model results
summary(Model_class_standard)
AIC(Model_class_standard)
BIC(Model_class_standard)
```

C.3. Student Model

```
    library(readxl)
library(Matrix)
library(lme4)
library(coda)
library(languageR)
library(lmerTest)
```

```
#Check the normality of the residuals
par(mfrow=c(1,2))
qqnorm(student_survey$MA, main = "MA")
qqnorm(log(student_survey$MA), main = "ln(MA)")

#Check normality of the explanatory variables
par(mfrow=c(2,4))
qqnorm(student_survey$leadership, main = "Leadership")
qqnorm(student_survey$helping, main = "Helping/Friendly")
qqnorm(student_survey$understanding, main = "Understanding")
qqnorm(student_survey$freedom, main = "Freedom/Student␣resp.")
qqnorm(student_survey$uncertain, main = "Uncertain")
qqnorm(student_survey$dissatisfied, main = "Dissatisfied")
qqnorm(student_survey$admonishing, main = "Admonishing")
qqnorm(student_survey$strict, main = "Strict")

boxplot(student_survey$leadership, main = "Leadership")
boxplot(student_survey$helping, main = "Helping/Friendly")
boxplot(student_survey$understanding, main = "Understanding")
boxplot(student_survey$freedom, main = "Freedom/Student␣resp.")
boxplot(student_survey$uncertain, main = "Uncertain")
boxplot(student_survey$dissatisfied, main = "Dissatisfied")
boxplot(student_survey$admonishing, main = "Admonishing")
boxplot(student_survey$strict, main = "Strict")

#Check correlation between the explanatory variables
cor(student_survey$leadership, student_survey$helping)
cor(student_survey$leadership, student_survey$understanding)
cor(student_survey$leadership, student_survey$freedom)
cor(student_survey$leadership, student_survey$uncertain)
cor(student_survey$leadership, student_survey$dissatisfied)
cor(student_survey$leadership, student_survey$admonishing)
cor(student_survey$leadership, student_survey$strict)

cor(student_survey$helping, student_survey$understanding)
cor(student_survey$helping, student_survey$freedom)
cor(student_survey$helping, student_survey$uncertain)
cor(student_survey$helping, student_survey$dissatisfied)
cor(student_survey$helping, student_survey$admonishing)
cor(student_survey$helping, student_survey$strict)

cor(student_survey$understanding, student_survey$freedom)
cor(student_survey$understanding, student_survey$uncertain)
cor(student_survey$understanding, student_survey$dissatisfied)
cor(student_survey$understanding, student_survey$admonishing)
cor(student_survey$understanding, student_survey$strict)

cor(student_survey$freedom, student_survey$uncertain)
cor(student_survey$freedom, student_survey$dissatisfied)
cor(student_survey$freedom, student_survey$admonishing)
cor(student_survey$freedom, student_survey$strict)

cor(student_survey$uncertain, student_survey$dissatisfied)
cor(student_survey$uncertain, student_survey$admonishing)
cor(student_survey$uncertain, student_survey$strict)
```

```
cor(student_survey$dissatisfied, student_survey$admonishing)
cor(student_survey$dissatisfied, student_survey$strict)

cor(student_survey$admonishing, student_survey$strict)

#Create the standard linear model
Model_student_standard <- lm(log(MA) ~ leadership + helping + understanding
                              + freedom + uncertain + dissatisfied + admonishing
                              + strict, data = student_survey)

#Print model results of the standard linear model
summary(Model_student_standard)
AIC(Model_student_standard)
BIC(Model_student_standard)

#Create the three-level mixed-effects model
Model_student_three <- lmer(log(MA) ~ leadership + helping + understanding
                             + freedom + uncertain + dissatisfied + admonishing
                             + strict + (leadership + helping + understanding
                                         + freedom + uncertain + dissatisfied
                                         + admonishing
                                         + strict|Class_ID/Teacher_ID)
                             , data = student_survey)

#Print model results of the three-level mixed-effects model
summary(Model_student_three)
r.squaredGLMM(Model_student_three)
AIC(Model_student_three)
BIC(Model_student_three)

#Create the Student-Class model
Model_student-class <- lmer(log(MA) ~ leadership + helping + understanding
                             + freedom + uncertain + dissatisfied + admonishing
                             + strict + (leadership + helping + understanding
                                         + freedom + uncertain + dissatisfied
                                         + admonishing
                                         + strict|Class_ID)
                             , data = student_survey)

#Print model results of the Student-Class model
summary(Model_student-class)
r.squaredGLMM(Model_student-class)
AIC(Model_student-class)
BIC(Model_student-class)

#Create the Student-Teacher model
Model_student-teacher <- lmer(log(MA) ~ leadership + helping + understanding
                             + freedom + uncertain + dissatisfied + admonishing
                             + strict + (leadership + helping + understanding
                                         + freedom + uncertain + dissatisfied
                                         + admonishing
                                         + strict|Teacher_ID)
                             , data = student_survey)

#Print model results of the Student-Teacher model
summary(Model_student-teacher)
```

```
r.squaredGLMM(Model_student-teacher)
AIC(Model_student-teacher)
BIC(Model_student-teacher)
```