VILNIUS UNIVERSITY
FACULTY OF MATHEMATICS AND INFORMATICS
MODELLING AND DATA ANALYSIS
MASTER'S STUDY PROGRAMME

Factors Affecting Students' Academic Achievement: Multilevel Structural Equation Model

Master's thesis

Author: Alina Alencenovič
VU email address: alina.alencenovic@mif.stud.vu.lt
Supervisor: Prof. Habil. Dr. Vydas Čekanavičius

Vilnius

2024

Factors Affecting Students' Academic Achievement: Multilevel Structural Equation Model

Abstract

The purpose of this thesis is to investigate the factors that affect academic achievement among students in Lithuania using data from the Program for International Student Assessment (PISA) 2018. Multilevel structural equation modeling (MSEM) is employed to analyze the impact of student and school level characteristics on academic achievement. The results obtained at the student level highlight the significance of factors such as parental support, bullying experiences, language spoken at home, and socioeconomic status. Self-efficacy is identified as a mediating factor affecting academic achievement. At the school level, factors such as school type, size, student-teacher ratio, and school socioeconomic status are found to have a significant impact on the average student's academic performance. The findings, attempts at alternative models, and prospects for future study are further discussed.

Keywords: Multilevel structural equation modeling, multilevel modeling, student achievement, PISA, categorical data.

Mokinių akademinius pasiekimus lemiantys veiksniai: daugiapakopis struktūrinių lygčių modelis

Santrauka

Šio darbo tikslas yra ištirti veiksnių, lemiančių Lietuvos mokinių akademinius pasiekimus, įtaką pasitelkus duomenis iš 2018m. tarptautinio penkiolikmečių tyrimo (PISA). Darbe taikomas daugiapakopis struktūrinių lygčių modelis (MSEM) siekiant išanalizuoti skirtingų veiksnių poveikį akademiniams pasiekimams mokinio ir mokyklos lygmenyse. Gauti rezultatai nurodo, kad mokinio lygmenyje tėvų palaikymas, mokykloje patiriamos patyčios, namuose vartojama kalba ir socialinė-ekonominė padėtis turi reikšmingą įtaką mokinių pasiekimams. Saviveiksmingumas pabrėžiamas kaip tarpininkaujantis veiksnys. Mokyklos lygmenyje nustatyta, kad tokie veiksniai kaip mokyklos tipas, dydis, mokytojų ir moksleivių skaičiaus santykis ir mokyklos socialinė-ekonominė padėtis turi reikšmingą poveikį vidutiniams mokinių akademiniams pasiekimams. Išsamiau aptariami gauti rezultatai, alternatyvių modelių taikymo bandymai ir perspektyvos tolimesniems darbams.

Raktiniai žodžiai: daugiapakopis struktūrinių lygčių modelis, daugiapakopis modeliavimas, mokinių pasiekimai, PISA, kategoriniai duomenys.

Contents

1 Introduction

Education plays a crucial role in advancing society as it shapes the intellectual, social, and economic landscapes. Numerous analyses have been conducted with a focus on school processes as analysts are interested in comprehending the quality of education, student achievement, and school outcomes.

According to the National Audit Office of Lithuania, there is insufficient progress in education in the country. Despite the increase in funding for education, the number of students and educational institutions is decreasing. Furthermore, there has been no significant reduction in the achievement gap between students attending schools in favorable and unfavorable social, economic, and cultural environments. This is due to the lack of measures to reduce the achievement gap and ensure consistently high-quality education. Moreover, many schools fail to monitor the progress of individual students, lack modern tools, and do not provide adequate support or implement inclusive practices in education. The results of international student achievement assessments also indicate no overall improvement in Lithuanian students' achievements [19].

Since 2006, Lithuania has been participating in the Programme for International Student Assessment (PISA) organized by the Organisation for Economic Cooperation and Development (OECD) to measure the academic achievements of its students in mathematics, science, and reading. However, according to the 2018 PISA report, Lithuania's students have shown no significant improvement in their scores in these subjects over the years. In fact, in all three subjects, Lithuanian students scored below the average score of the participating OECD countries, maintaining a consistent gap between Lithuania's and the OECD's scores [31].

As the field of education is facing various obstacles that require consideration, the differences in the quality of education and the complex nature of students' experiences indicate the need for detailed research. Therefore, it is important to have a deeper understanding of the factors that affect students' academic performance. How do personal qualities affect the student's academic achievement? How does the school environment impact students' educational experiences?

The purpose of this thesis is to investigate various factors that affect students' academic performance by utilizing multilevel structural equation modeling (MSEM) and analyzing both student-level characteristics and school-level variables using PISA 2018 data. The thesis begins with a literature review that summarizes previous studies conducted in different countries on student and school factors that impact student achievement. Next, the theoretical concepts of multilevel structural equation modeling are presented. Prior to empirical analysis, the methodology and suggestions for working with PISA data are discussed. Finally, the data is analyzed and an MSEM model is developed to investigate the factors that influence academic achievement. Lastly, a discussion of the results, reflection on the limitations of the study, and suggestions for future research are presented.

2 Literature review

2.1 Parental support

Given that students spend a substantial amount of time outside the school, parental support has emerged as a crucial element influencing students' academic performance in educational research. The term "parental support" refers to the actions and activities parents take to support and improve children's learning and academic performance at home and school [9]. A survey carried out in Lithuania has shown that parental involvement in school life is associated with children's achievements: parents of high-achieving students are more likely to engage in school activities than parents of low-achieving students [7]. Additionally, active parental engagement in the educational process influenced students' attitudes toward learning, their learning motivation, and increased academic achievements [20]. Out-of-school parental support is usually understood as emotional support perceived by students at home. The emotional support of parents and other adult family members has been shown to enhance self-confidence, ability to overcome challenges, capacity to address issues that arise in school, and improve academic achievements ([2], [28]).

The topic of parental support has been studied in various countries. In Turkey, a study was conducted using the PISA 2018 data, where a multilevel structural equation modeling approach was used to investigate the relationship between emotional parental support and students' reading achievement [6]. The results showed a positive impact of parental support on students' reading achievement both at an individual (student) level and across schools. Similarly, another study using multilevel modeling approach was conducted in South Korea, Turkey, and the USA to analyze the effect of parental support on science achievements at the student level, and the results showed a positive effect [5].

2.2 Experience being bullied

The issue of bullying is a concern that affects people all around the world. Bullying is defined as intentional harmful behavior, characterized by repeated acts of aggression with a desire to cause harm and control psychologically or physically in interpersonal relationships. According to a survey conducted in Lithuania, the detrimental effects of bullying at school include persistent tension and stress, conflicts with peers, behavioral changes, decreased motivation to learn, and academic failure [40]. Those who do not experience bullying tend to have better academic performance, improved mental health and well-being and are less likely to miss school [11].

A study employing the MSEM method to research the extent to which bullying affects reading achievements on student and school levels was conducted in Canada [12]. The findings revealed that bullying negatively affected not only reading achievement but also reading self-efficacy at both student and school levels. In addition, another study performed in Canada showed that bullying had a negative effect on mathematics achievement at the school level [22]. However, a study conducted in Chile proposed that bullying significantly negatively affects mathematics achievements on the student level only [3].

2.3 Self-efficacy

Self-efficacy refers to an individual's belief in one's ability to successfully perform a task or achieve a specific goal. It refers to an individual's confidence in their capacity to carry out a necessary course of action in order to accomplish desired results. Self-efficacy is a significant motivator for performance since it determines how challenging a task is perceived, which in turn affects the amount of effort and persistence invested in completing it [41].

A study was conducted in the USA to understand the relationship between self-efficacy and students' science achievements. The study was based on MSEM and used Trends in International Mathematics and Science Study (TIMSS) 2011 data. The research findings revealed that students' self-efficacy has a significant positive impact on science achievement [13]. In another study carried out in China, it was observed that self-efficacy positively influences students' engagement in foreign language classes. However, there was no direct significant effect on their foreign language achievements [43]. This suggests that self-efficacy can help students perform better in a subject, but its effects may vary depending on the discipline.

2.4 Socioeconomic status

Socioeconomic status (SES) has always been considered an important factor in educational research, as it provides a comprehensive view of the various economic, social, and cultural factors that influence academic achievement. In Lithuania, students who come from a more advantageous socioeconomic background tend to have higher motivation to learn and achieve better academic results compared to students from less favorable socioeconomic backgrounds. Research shows that students who lack achievement motivation tend to perceive the classroom environment as less motivating, less organized, and less predictable [10].

Analysis of PISA 2018 results in Lithuania revealed that since 2006, when Lithuania first started participating in assessment, a significant achievement gap between students of different socioeconomic strata remained, showing that students with higher SES often perform better academically [45].

The impact of socioeconomic status on academic achievement is often studied alongside other variables. The same study conducted in Turkey has shown a significant positive effect of SES on reading achievement at student and school levels together with its mediating effect on teachers' hindering, indicating that families belonging to higher SES are more likely to have the ability to influence teachers to exhibit fewer hindering behaviors ([6].

3 Theoretical framework

3.1 Multilevel structural equation modeling

When data is collected at multiple levels, it forms a hierarchical structure where observations at one level are nested within higher levels. For instance, in social research, individuals (level 1) are nested into groups (level 2). This structure has two crucial implications. Firstly, level 1 observations are often not independent, and ignoring this fact can lead to biased estimates. Secondly, single-level analyses that overlook this structure can result in misleading outcomes, particularly when group-level results are interpreted at the individual level, and vice versa. [4].

Multilevel modeling (MLM) is a statistical method used to analyze nested data. This technique allows for the variability of the outcome to be modeled and explained by higher-level predictors. By doing so, MLM helps to identify significant differences between groups and provides insights into the sources of variability. However, the typical MLM has its drawbacks. These include the limitation to a single level 1 outcome variable and the use of only observed variables in the analysis ([14], [4]).

When the variables of interest cannot be measured directly, structural equation modeling (SEM) is employed. Instead of observed variables, the latent variables are interpreted as constructs that lie beneath the measured items and cause dependence between them [36]. The relationships between latent and observed variables are outlined in the measurement model and the structural model specifies the directional relationships between latent variables and other observed variables not in the measurement model.

Multilevel structural equation modeling (MSEM) is a statistical method that combines the benefits of SEM with multilevel models. It expands the measurement and structural models of SEM to incorporate random effects that are compatible with MLM. This allows for the use of several indicators for every latent variable, thereby addressing measurement error. Additionally, it enables the modeling of multiple outcome variables at various levels of analysis simultaneously. Moreover, complex relationships between variables of interest can be defined and calculated simultaneously, which is frequently the case in models with mediation [27].

The main idea behind MSEM is that a model can be constructed by splitting the total covariance matrix into separate within-group (level 1) and between-group (level 2) covariance matrices by latent decomposition [27]. In this method, individual observations, denoted as y_{ij}, are broken down into two components: the mean of group j, represented as y_j, and the individual's deviation from the mean, represented as y_{ij}:

$$y_{ij} = y_j + (y_{ij} - y_j) \tag{1}$$

Following this decomposition, since y_{ij} and y_j are uncorrelated, the total covariance structure is divided into the separate within and between covariance matrices:

$$\Sigma_T = \Sigma_W + \Sigma_B. \tag{2}$$

As the MSEM model is an extension of the SEM model that incorporates a multilevel aspect, it can be defined by the SEM composition of measurement and structural models. The within-level random intercepts serve as between-level latent variables and capture the variability in the means of the observed and latent variables. The equations at the within-level are formulated as follows:

$$y_{ij} = \nu_j + \Lambda_W \eta_{ij} + \varepsilon_{Wij}, \tag{3}$$

$$\eta_{ij} = \mu_j + B_W \eta_{ij} + \zeta_{Wij}. \tag{4}$$

Similarly, the between-level equations are written as:

$$y_j = \nu + \Lambda_B \eta_j + \varepsilon_{Bj}, \tag{5}$$

$$\eta_j = \mu + B_B \eta_j + \zeta_{Bj}, \tag{6}$$

where y_{ij} is a vector of within-level observed variables, y_j is a vector representing the group means of the latent decomposition, ν_j and ν are vectors of intercepts, Λ_W and Λ_B are within and between level factor loadings matrices, ε_{Wij} and ε_{Bj} are vectors of within and between level residuals, η_{ij} and η_j are vectors of within and between level latent variables, μ_j and μ are vectors of intercepts, B_W and

B_B are matrices of within and between level regression coefficients, ζ_{Wij} and ζ_{Bj} are vectors of within and between level residual variances. Equations (3) and (5) represent the measurement models liking observed variables to underlying factors at each level, and equations (4) and (6) represent structural models of relationships between latent variables at each level ([26], [27]).

In order to clarify the conceptual framework described by the equations, Figure 1 illustrates a hypothetical multilevel structural equation model. At level 1, measurement pathways connect observed variables (represented by rectangles) to latent constructs (represented by circles). Equations (7) and (8) describe the measurement and structural models of Level 1, respectively.

At level 2, the structural paths explain the relationships among the group mean latent variables and add the intercept component (represented by a triangle). Equations (9) and (10) describe the measurement and structural models of Level 2. To prevent visual overload, the residuals of observed and latent variables are left out of the graphical depiction.

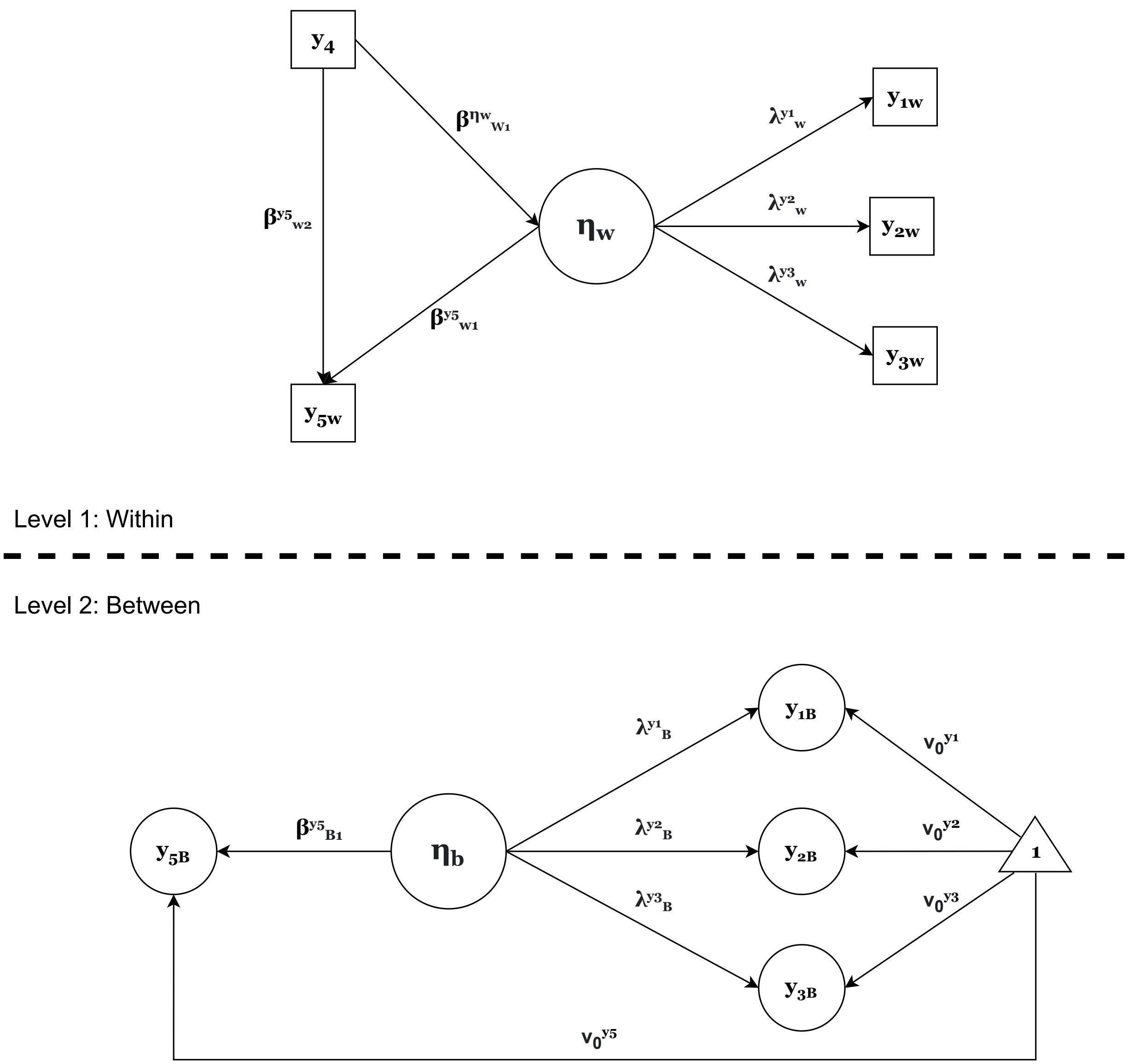

Figure 1: Illustration of a MSEM model.

Level 1 measurement model:

$$\begin{cases} y_{1Wij} = \nu_{0j}^{y_1} + \lambda_W^{y_1}\eta_{Wij} + \varepsilon_{ij}^{y_1}, \\ y_{2Wij} = \nu_{0j}^{y_2} + \lambda_W^{y_2}\eta_{Wij} + \varepsilon_{ij}^{y_2}, \\ y_{3Wij} = \nu_{0j}^{y_3} + \lambda_W^{y_3}\eta_{Wij} + \varepsilon_{ij}^{y_3}. \end{cases} \tag{7}$$

Level 1 structural model:

$$\begin{cases} \eta_{Wij} = \mu_{0j}^{\eta_W} + \beta_{W1}^{\eta_W} y_{4ij} + \zeta_{ij}^{\eta_W}, \\ y_{5Wij} = \nu_{0j}^{y_5} + \beta_{W1}^{y_5}\eta_{Wij} + \beta_{W2}^{y_5} y_{4ij} + \varepsilon_{ij}^{y_5}. \end{cases} \tag{8}$$

Level 2 measurement model:

$$\begin{cases} y_{1Bj} = \nu_{0j}^{y_1} = \nu_0^{y_1} + \lambda_B^{y_1}\eta_{Bj} + \varepsilon_j^{y_1}, \\ y_{2Bj} = \nu_{0j}^{y_2} = \nu_0^{y_2} + \lambda_B^{y_2}\eta_{Bj} + \varepsilon_j^{y_2}, \\ y_{3Bj} = \nu_{0j}^{y_3} = \nu_0^{y_3} + \lambda_B^{y_3}\eta_{Bj} + \varepsilon_j^{y_3}. \end{cases} \tag{9}$$

Level 2 structural model:

$$y_{5Bj} = \nu_{0j}^{y_5} = \nu_0^{y_5} + \beta_{B1}^{y_5}\eta_{Bj} + \varepsilon_j^{y_5}. \tag{10}$$

The representation style of equations (7)-(10) is adapted from the book "Multilevel Structural Equation Modeling" by Bruno Castanho Silva, Constantin Manuel Bosancianu, and Levente Littvay [38] and will be applied throughout the analysis.

3.2 Sample size

In MSEM, the size of the sample becomes an important consideration, particularly at level 2 and higher. Studies have shown that for simple models, having a level 2 sample size of 20 groups is usually adequate to obtain unbiased results in simple models. However, others suggest a general rule of thumb, known as the 30/30 rule, which recommends a minimum of 30 groups, each with 30 observations if the model is relatively simple. If cross-level interactions are included, a 50/20 rule might be preferable. Although some have argued that this rule could go up to 100/10 [38]. However, fixing the size of level 1 to a certain amount is not as important because the model does not assume equal sample sizes. At level 1, small sample sizes do not necessarily pose an issue as long as there are larger groupings. However, small groups can have drawbacks, such as the inability to identify the model as it becomes more complex. Therefore, a recommended minimum level 1 sample size of 10 observations per group is advised [16].

When working with categorical data, it is recommended to have a sufficiently large sample size to ensure accurate and reliable results. To minimize the standard error of parameter estimates, it is suggested to have a minimum of 200 groups with at least 10 observations in each group. Therefore, the sample size of at least 200 groups with at least 10 observations in each group is to be used [17].

3.3 Fit indices

In MSEM, evaluating the proposed model and its compatibility with the data is just as important as it is in single-level SEM. The same test statistics and fit indices that are used to evaluate model fit in single-level SEM are also employed in MSEM, as the entire multilevel model is evaluated simultaneously. However, it's worth noting that the standard approach for assessing model fit may be dominated by the fit at the lower level, since the sample size of multilevel data is usually much larger at the lower level than at the higher level. Therefore, it may not be as sensitive to detect a lack of fit at the higher level. To assess the model fit in MSEM, a number of fit indices are used, including RMSEA, CFI, TLI, and SRMR.

- Root mean square error of approximation (RMSEA):

  $$RMSEA = \sqrt{max\left(0, \frac{\chi^2 - df}{df(N-1)}\right)},$$

  where χ^2 is the chi-square test statistic of the tested model with df degrees of freedom.

  RMSEA is a metric that measures how well a statistical model fits the data. A smaller value of RMSEA indicates a better overall fit of the model to the data, with 0 indicating a perfect fit. A large value of RMSEA indicates a misspecification in either the within or between levels of the model. To determine if a model is well-fitting, an RMSEA value of 0.05 or less is required.

- Compararive fit index (CFI)

  $$CFI = 1 - \frac{max\left[\left(\chi^2_{Hypothesized} - df_{Hypothesized}\right), 0\right]}{max\left[\left(\chi^2_{Baseline} - df_{Baseline}\right), \left(\chi^2_{Hypothesized} - df_{Hypothesized}\right), 0\right]},$$

  where $\chi^2_{Hypothesized}$ is the chi-square test statistic of the tested model with $df_{Hypothesized}$ degrees of freedom and $\chi^2_{Baseline}$ is the chi-square test statistic of the baseline model with $df_{Baseline}$ degrees of freedom.

  CFI is a statistical measure used to determine how well a proposed model fits the data in comparison to a baseline model in which all variables are uncorrelated. Similar to RMSEA, CFI is a way to assess the overall goodness of fit of a multilevel model. A higher CFI score denotes a better model fit to the data. Generally, a CFI of 0.95 or higher is considered an indication of a good fit model.

- Tucker-Lewis Index (TLI)

  $$TLI = \frac{\chi^2_{Baseline}/df_{Baseline} - \chi^2_{Hypothesized}/df_{Hypothesized}}{\chi^2_{Baseline}/df_{Baseline} - 1},$$

  where again, $\chi^2_{Hypothesized}$ is the chi-square test statistic of the tested model with $df_{Hypothesized}$ degrees of freedom and $\chi^2_{Baseline}$ is the chi-square test statistic of the baseline model with $df_{Baseline}$ degrees of freedom.

TLI is a global fit index, much like CFI. A higher TLI value suggests a better fit of the model to the data. A good fit model is indicated by a TLI of 0.95 or more.

- Standardized root mean squared residual (SRMR)

$$SRMR = \sqrt{\frac{2\sum_{k=1}^{p}\sum_{j=1}^{k}\left[\frac{s_{kj}-\hat{\sigma}_{kj}}{s_{kk}s_{jj}}\right]^2}{p(p+1)}},$$

where s_{kj} represents the sample covariance between variables k and j, while $\hat{\sigma}_{kj}$ denotes the corresponding model-implied covariance between these variables.s_{kk} and s_{jj} are the sample standard deviations for variables k and j respectively, and p indicates the total number of variables in the model being analyzed.

SRMR assesses the average magnitude of the differences between observed and expected correlations $s_{kj} - \hat{\sigma}_{kj}$ as a measure of model fit criterion. A lower value of SRMR indicates a better fit and a value of 0.08 or less is generally considered to be a good fit.

Since in MSEM the covariance matrices for the within and between levels are calculated separately, SRMR can also be calculated separately for the within level (SRMR_W) and the between level (SRMR_B). Currently, this is the only fit index that is provided for each level and reported in the statistical modeling program *Mplus*.

In MSEM, the reporting of the χ^2 fit indice is often excluded as compared to single-Level SEM. The χ^2 test can only indicate if the model-implied covariance matrix is equal to the observed covariance matrix. It cannot provide any other information about the magnitude or location of the difference or mismatch [18]. Hence, the χ^2 statistic is usually not included in MSEM model evaluation. However, it still plays a significant role in generating other fit statistics, as described earlier.

3.4 Estimation with categorical data

A common challenge in applying various statistical models is determining how to handle variables measured on a categorical scale. Numerous researches argue whether categorical variables can be handled as if they were continuous, or whether they need to be substituted with genuinely continuous variables in causal models [44]. While some suggest that if the number of categories of the variable exceeds 4, it may be deemed as continuous, others oppose stating that such methods introduce bias and undermine the parameter estimates of a model ([37], [42]). Therefore, a threshold model can be implied where for each categorical variable, it is assumed that there exists a corresponding underlying continuous variable:

$$y = \begin{cases} 1, & \text{if } -\infty < y^* \leq \tau_1, \\ 2, & \text{if } \tau_1 < y^* \leq \tau_2, \\ 3, & \text{if } \tau_2 < y^* \leq \tau_3, \\ \dots & \\ k, & \text{if } \tau_{k-1} < y^* \leq \infty, \end{cases} \tag{11}$$

where y is an observed categorical variable, y^* is a latent continuous variable and $\tau_1, \tau_2, \ldots, \tau_{k-1}$ are the thresholds.

It is assumed that y^* follows a standard normal distribution with mean 0 and variance 1 and the thresholds $\tau_1, \tau_2, \ldots, \tau_{k-1}$ are unknown parameters. The likelihood of observing a specific response category is determined by the cumulative probability of the latent variable landing within the corresponding threshold interval [21]. For example, when a variable with 3 categories is observed, the probability of observing a response of 1 is given by $P(-\infty < y^* \leq \tau_1) = \Phi(\tau_1)$, where Φ is the cumulative distribution function of the standard normal distribution. The probability of observing a response of 2 is then given by $P(\tau_1 < y^* \leq \tau_2) = \Phi(\tau_2) - \Phi(\tau_1)$ (Figure 2).

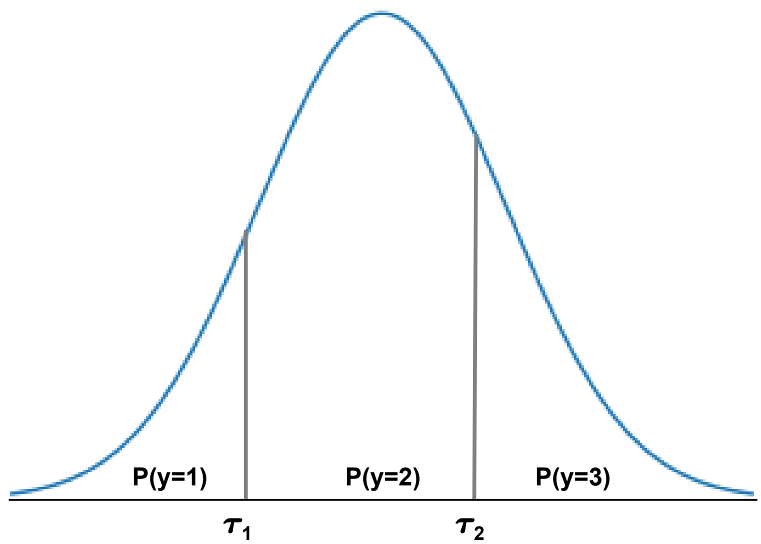

Figure 2: Example of a variable with 3 categories.

The thresholds can be estimated by inversing the cumulative distribution function of the standard normal distribution. Then the probability of a response in category i is

$$\pi_i = \mathbb{P}(y = i) = \mathbb{P}(\tau_{i-1} < y^* \leq \tau_i) = \int_{\tau_{i-1}}^{\tau_i} \phi(u)du = \Phi(\tau_i) - \Phi(\tau_{i-1}), \tag{12}$$

therefore

$$\tau_i = \Phi^{-1}(\pi_1 + \pi_2 + \ldots + \pi_i), \quad i = 1, 2, \ldots, k-1. \tag{13}$$

where Φ^{-1} represents the inverse function of the standard normal distribution. The term $(\pi_1 + \pi_2 + \ldots + \pi_i)$ signifies the probability of a response falling into category i or lower. Since the probabilities π_i are unknown quantities, they can be estimated by the corresponding proportion p_i of responses in category i:

$$\hat{\tau}_i = \Phi^{-1}(p_1 + p_2 + \ldots + p_i), \quad i = 1, 2, \ldots, k-1, \tag{14}$$

where the term $(p_1 + p_2 + \ldots + p_i)$ is the proportion of response cases in category i or lower in the sample [21].

The obtained threshold values, together with the polychoric correlations, which estimate the correlation between pairs of hypothesized normally distributed underlying continuous variables of observed categorical variables, are then employed in the parameter estimation process. Diagonally weighted least squares (DWLS) estimator is preferred when constructing an MSEM model with categorical data, where polychoric correlations and obtained thresholds are used:

$$F_{DWLS} = [s - \sigma(\theta)]' W_D^{-1} [s - \sigma(\theta)], \tag{15}$$

where θ represents the vector of model parameters, W_D is the diagonal weight matrix, $\sigma(\theta)$ is the model-implied vector containing nonredundant elements of a model-implied covariance matrix, and s is the vector containing the elements of threshold and polychoric correlation estimates [24].

The DWLS estimator is also useful when working with data that is not normally distributed since the estimator does not rely on assumptions of normality or continuity. This method is designed to provide accurate parameter estimates, even when the data is skewed, the sample size is small, or a large model is estimated [29].

4 PISA methodology

4.1 Introduction to PISA

Initiated by the Organisation for Economic Cooperation and Development (OECD), the Programme for International Student Assessment (PISA) aims to assess and compare the knowledge and skills of 15-year-old students across participating countries. Launched in the year 2000 it has become a triennial benchmark that provides a comprehensive view of the effectiveness of education systems and their ability to prepare students for real-world challenges. PISA is globally recognized and in 2018, 79 countries and economies participated in PISA.

The PISA assessment evaluates students' knowledge and skills in the core school subjects of reading, mathematics, and science. In every round of PISA, one of the core subjects is tested in detail, and the main area of focus in 2018 was reading. Through computer-based tests, distributed to students and school principals, PISA gathers information about students and their family backgrounds, including their economic, social, and cultural capital, attitudes toward learning, and experiences both in and outside of school. It also examines aspects of schools, such as the quality of resources, management, staffing practices, curricular and extracurricular activities offered, class size, classroom and school climate, and school type and size.

Policymakers around the world rely on the results of PISA surveys to evaluate the knowledge and abilities of students in their own country compared to those in other participating countries. This allows them to gain an understanding of the strengths and weaknesses of their own education systems and to identify areas that require improvement [32].

4.2 Sampling design and weights

The sampling process for PISA follows a detailed procedure to select a representative group of 15-year-old students from each participating country. PISA uses a multistage, stratified sampling design, which begins with selecting schools within each country, taking into account aspects such as size and location. After that, a sample of students is chosen from these schools with equal probabilities. Each sampled student is assigned a final weight that indicates the number of students from the population they represent [33].

The final weight W_{ij} for student i in school j consists of two base weights (school base weight and the within-school base weight) and four adjustment factors:

$$W_{ij} = w_{1j} \cdot t_{1j} \cdot f_{1j} \cdot w_{2ij} \cdot f_{2ij} \cdot t_{2ij}, \tag{16}$$

where w_{1j} is the school base weight - reciprocal of the probability of the school j being included in the sample;

t_{1j} is a trimming factor for school base weight that is intended to reduce unexpectedly large values of w_{1j};

f_{1j} is an adjustment factor to compensate for the non-participation of schools similar to school j, not compensated by replacement schools.

w_{2ij} is the within-school base weight, calculated by taking the reciprocal of the probability of selecting a student i from the selected school j;

f_{2ij} is an adjustment factor that compensates for non-participation by students in the same school, non-response cell, explicit stratum and, if sample size permits, high/low grade and gender categories.

t_{2ij} is the final student weight trimming factor used to reduce the weights of students who have exceptionally large values for the product of all the preceding weight components. [35].

However, the PISA provided weights are appropriate for analyzing either the student or school level separately. When analyzing student and school levels simultaneously (such as in MSEM) the weights have to be scaled appropriately to account for the hierarchical data structure and eliminate bias in the estimates ([25], [39]). For the student level, final student weights are scaled using an Ecluster method:

$$w_{ij}^* = w_{ij} \frac{n_j^*}{\sum_i w_{ij}}, \tag{17}$$

where n_j^* is the effective sample size for cluster j:

$$n_j^* = \frac{(\sum_i w_{ij})^2}{\sum_i w_{ij}^2}. \tag{18}$$

For the school level, final school weights are used without transformation.

4.3 Plausible values

The main objective of the PISA assessment is to measure students' cognitive abilities and provide summary statistics on a population level within each country. The evaluation process involves administering a set of questions or items within a specific time frame on subjects like reading, mathematics, and science to students within the sampled schools. These assessments rely on a complex psychometric design that divides the test items into groups due to the large amount of material. As a result, each student only answers a fraction of the total assessment, leading to measurement error. To address this issue, plausible value methods are used to estimate individual proficiency. Plausible values can be defined as random values drawn from the posterior distributions of proficiency. Therefore, plausible values should not be considered as test scores [32].

For secondary analysis purposes, PISA datasets contain sets of 10 plausible values for each of the subject domains. When dealing with plausible values in the secondary analyses of students' academic performance, it is common to find two different shortcuts in some empirical studies. First, researchers frequently decide to use only one of the plausible values. Second, researchers use a shortcut to estimate student performance by averaging the plausible values. The standard errors of the relevant statistics are severely underestimated when using these options, leading to inaccurate results ([1], [23]). Therefore, to ensure the correct estimation during the research, analyses should be performed 10 times, once for each plausible value. Aggregated value statistics are then reported as final results [34].

5 Empirical application

5.1 Data overview

The PISA educational assessment data for secondary analysis includes datasets for students, teachers, and schools, each containing a vast number of variables. For this study, student and school datasets were used. Selected variables were deemed important to student educational outcomes in Lithuania and were used in the following analyses. The variable names used in the modeling part are provided together with their corresponding labels in the PISA datasets (denoted in square brackets). Additionally, the constructed latent variables are also described.

5.1.1 Student level items

- Parents' support

  Parents' support is a latent variable reflecting the supportive actions parents display for their children's education. It is measured by 3 student questionnaire items:

  PS1: "My parents support my educational efforts and achievements" [ST123Q02NA],

  PS2: "My parents support me when I am facing difficulties at school" [ST123Q03NA],

  PS3: "My parents encourage me to be confident" [ST123Q04NA].

  Items are rated on a 4-point Likert scale, ranging from 1 (strongly disagree) to 4 (strongly agree).

- Experience being bullied

  Experience being bullied is a latent variable representing the underlying construct of negative encounters with harassment, intimidation, or acts of interpersonal aggressiveness in the school environment. It is measured by 6 student questionnaire items:

BULLY1: "Other students left me out of things on purpose" [ST038Q03NA],

BULLY2: "Other students made fun of me" [ST038Q04NA],

BULLY3: "I was threatened by other students" [ST038Q05NA],

BULLY4: "Other students took away or destroyed things that belonged to me" [ST038Q06NA], BULLY5: "I got hit or pushed around by other students" [ST038Q07NA],

BULLY6: "Other students spread nasty rumors about me" [ST038Q08NA].

Items are rated on a 4-point Likert scale, ranging from 1 (never or almost never) to 4 (once a week or more).

- Self-efficacy
  Self-efficacy is a latent variable reflecting students' beliefs and confidence in their ability to handle various challenging situations and achieve desired results. It is measured by 4 student questionnaire items:

SEFF1: "I usually manage one way or another" [ST188Q01HA],

SEFF2: "I feel that I can handle many things at a time" [ST188Q03HA],

SEFF3: "My belief in myself gets me through hard times" [ST188Q06HA],

SEFF4: "When I'm in a difficult situation, I can usually find my way out of it" [ST188Q07HA].

Items are rated on a 4-point Likert scale, ranging from 1 (strongly disagree) to 4 (strongly agree).

- Academic achievements
  Academic achievement is a latent variable representing students' academic accomplishments between different academic disciplines. It is measured by 3 plausible values from the student dataset:

MATH Plausible value in mathematics [PV1-PV10MATH],

SCIENCE Plausible value in science [PV1-PV10SCIE],

READ Plausible value in reading [PV1-PV10READ],

As mentioned, plausible values are random values derived from the posterior distributions of proficiency.

- Gender [ST004D01T]
  The gender of students is a variable with values of 1 for females and 2 for males. For research purposes, it is rescaled to a dummy variable with values of 0 for males and 1 for females.

- Language spoken at home [ST022Q01TA]
  Language spoken at home is a variable with values 1 for the language of the test (or lithuanian) and 2 for other language. For research purposes again, it is rescaled to a dummy variable with values 0 for the language of the test and 1 for the other language.

- SES or Socioeconomic status [ESCS]
  SES, represented by the PISA index of economic, social, and cultural status is an OECD derived variable based on three factors associated with family background: parents' highest level of education, parents' highest occupational status, and home possessions, including books in the home [32].

5.1.2 School level items

- School type [SCHLTYPE]
  School type is a variable that distinguishes between public and private schools. The values include 1 for private independent, 2 for private government-dependent, and 3 for public. For research purposes, it is aggregated to a dummy variable with values 0 for private (comprised of private independent and private government-dependent schools) and 1 for public schools.

- School size [SCHSIZE]
  School size represents the total number of students enrolled in a school.

- Student-teacher ratio [STRATIO]
  The student-teacher ratio represents the average number of students per teacher.

To determine whether data is normally distributed, the Shapiro-Wilk normality test was performed on each variable. All categorical variables and most interval scale variables were found to be non-normally distributed, except for the following: PV3MATH, PV7MATH, PV9MATH, PV10MATH, PV5SCIE, and PV9SCIE.

5.2 Exploratory data analysis

At the beginning of the research, exploratory analysis is performed to obtain a grasp of the data's patterns and potential connections. The sample for Lithuania included 6885 students from 362 schools. Descriptive statistics with weighted means of mathematics, science, and reading literacy scores of different grouping versions are presented in Table 1.

According to the results, the overall scores in mathematics, science, and reading literacy were 481.19, 482.07, and 475.87 respectively, confirming that Lithuanian students scored lower in mathematics, science, and reading than the OECD average (489, 489, and 487 respectively). A comparison of the academic achievements of male and female students revealed a small difference in mathematics (2 points) and science (6 points) achievements and a higher gap in reading achievement (39 points), with the female students having higher average scores.

As Lithuanian schools use three main languages for instruction, namely Lithuanian, Polish and Russian, the effect of the language spoken at home on academic achievement also became a topic of interest. Analysis of language spoken at home displayed that students who spoke the test language (most often Lithuanian) at home tended to score over 40 points more among all three subjects compared to those who spoke other languages.

The sample was categorized into 5 groups of equal size based on the socioeconomic status of students. Three groups were chosen to represent the bottom 20% (values up to the 20th percentile), middle 20% (values between the 40th and 60th percentiles), and top 20% (values above the 80th percentile) of students in different SES categories. A difference of over 100 points was found between the students at the bottom and top 20% SES categories in all three subjects.

An examination of academic achievements by school type indicated differences between public and private schools. On average, students in private schools scored over 60 points more in all three subjects than their counterparts in public schools.

Grouping	Number of students	Mean Math	Mean Science	Mean Read
Sample	6885	481.19	482.07	475.87
Gender: Male	3377	479.98	479.09	456.97
Gender: Female	3508	482.45	485.18	495.63
Language at home: Test	6313	484.03	485.36	479.12
Language at home: Other	474	440.14	434.33	431.26
SES: bottom 20%	1339	432.82	436.26	427.08
SES: middle 20%	1338	478.07	478.06	471.69
SES: top 20%	1339	529.63	528.94	524.40
School: Private	333	544.39	545.38	539.73
School: Public	6552	478.45	479.32	473.10

Table 1: Descriptive statistics of student academic achievements in Lithuania.

For modeling purposes, listwise deletion was performed on the initial sample, to handle missing data. As some schools were represented by less than 10 students, they were excluded to ensure the right sample size, as indicated in the sample size section. The final sample used for modeling consisted of 4550 students from 217 schools.

5.3 MSEM model

Prior to model specification and estimation, it is important to determine the intraclass correlation coefficient (ICC) to analyze the need for multilevel analysis. ICC index represents the proportion of variance in an observed variable found at the higher level, rather than the individual level and is calculated as follows:

$$ICC = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2}, \tag{19}$$

where σ_B^2 is the between level variance and σ_W^2 is the within level variance.

Since ICC estimates are variance ratios, index values vary from 0 to 1, where greater values indicate a larger proportion of between-level variance. A multilevel analysis is typically indicated by an ICC value larger than 0.05. When ICCs are smaller than 0.05, multilevel modeling may not be beneficial in practice. Yet, studies have shown that large samples may still be affected by ICCs as low as 0.01 [8].

To assess the need for multilevel analysis in this research, ICCs were calculated for the variables representing academic achievements - 10 plausible values of each subject. The ranges of ICCs, presented in Table 2, indicate the necessity of multilevel analysis as all ICCs are greater than 0.05.

	Plausible value in Math	Plausible value in Science	Plausible value in Reading
ICC range	0.250 - 0.278	0.269 - 0.290	0.292 - 0.307

Table 2: ICCs of plausible values.

Based on the performed exploratory analysis and studies described in the literature review, an MSEM model, displayed in Figure 3, is proposed. The model considers the potential mediating effect to assess the complexity of student achievement. As mentioned in the data overview, the variables *Parents' Support*, *Experience Being Bullied*, *Self-efficacy* and *Academic Achievement* are the latent variables with *General Self-efficacy* being the mediating factor. *Gender*, *Language at Home*, *Student-*

Teacher Ratio, *School Size* and *School Type* are the covariates and *Math*, *Science*, *Read*, *SES* are observed variables split into student and school level variables via latent decomposition, as described in the theoretical framework section. The indicators of latent variables and residuals are intentionally omitted from the graphical representation for clarity.

The model's graphical representation is also expressed in the set of equations (Equations (20)-(25)) which provides a more detailed understanding of the model. Here subscripts represent student i in school j. In equations, variable names are abbreviated without losing their meaning.

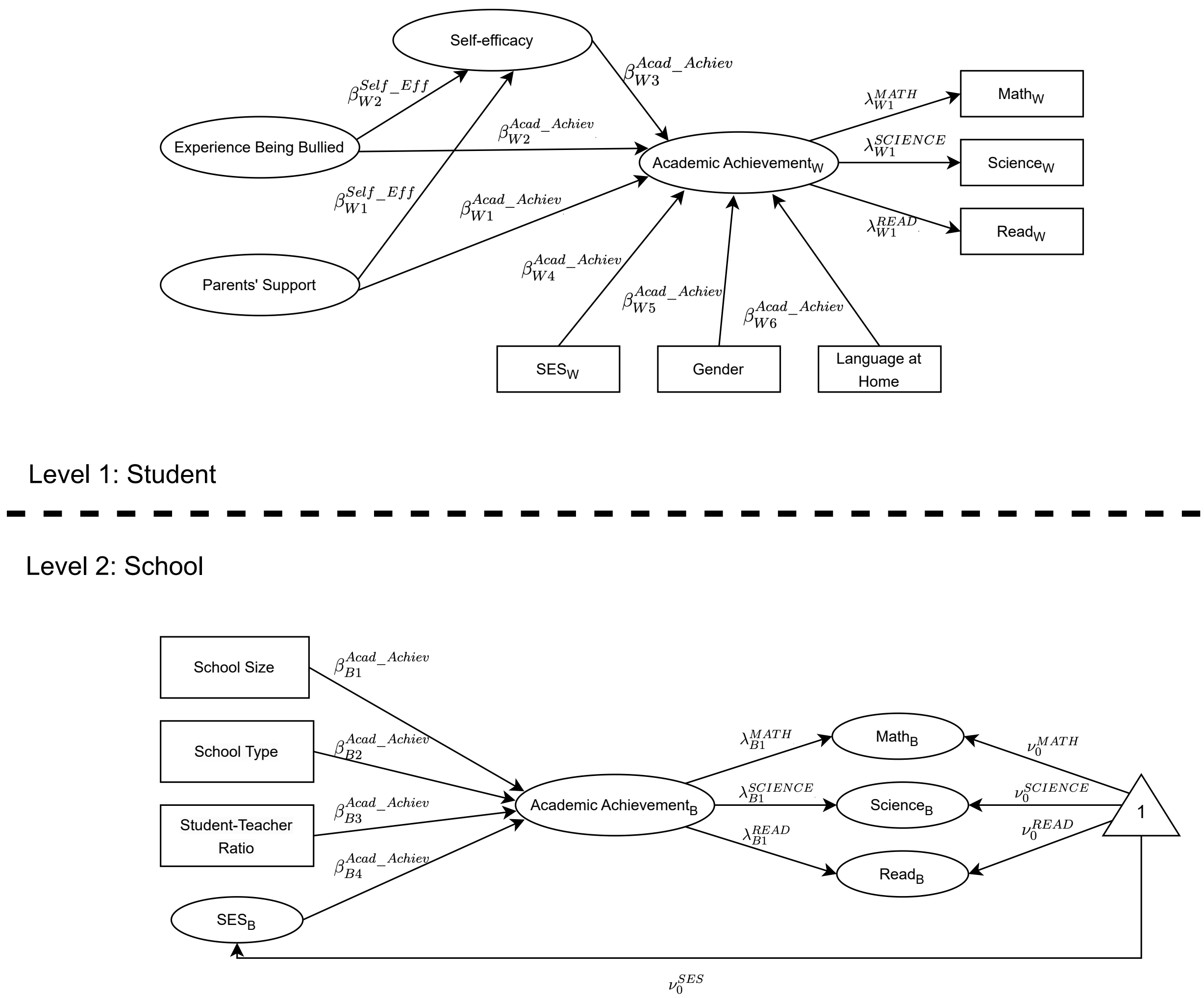

Figure 3: Hypothesized MSEM model.

Level 1 (student) measurement model:

$$
\begin{cases}
PS1_{ij} = \nu_0^{PS1} + \lambda_1^{PS1} Parents\_Support_{ij} + \varepsilon_{ij}^{PS1}, \\
PS2_{ij} = \nu_0^{PS2} + \lambda_1^{PS2} Parents\_Support_{ij} + \varepsilon_{ij}^{PS2}, \\
PS3_{ij} = \nu_0^{PS3} + \lambda_1^{PS3} Parents\_Support_{ij} + \varepsilon_{ij}^{PS3}, \\
BULLY1_{ij} = \nu_0^{BULLY1} + \lambda_1^{BULLY1} Being\_Bullied_{ij} + \varepsilon_{ij}^{BULLY1}, \\
BULLY2_{ij} = \nu_0^{BULLY2} + \lambda_1^{BULLY2} Being\_Bullied_{ij} + \varepsilon_{ij}^{BULLY2}, \\
BULLY3_{ij} = \nu_0^{BULLY3} + \lambda_1^{BULLY3} Being\_Bullied_{ij} + \varepsilon_{ij}^{BULLY3}, \\
BULLY4_{ij} = \nu_0^{BULLY4} + \lambda_1^{BULLY4} Being\_Bullied_{ij} + \varepsilon_{ij}^{BULLY4}, \\
BULLY5_{ij} = \nu_0^{BULLY5} + \lambda_1^{BULLY5} Being\_Bullied_{ij} + \varepsilon_{ij}^{BULLY5}, \\
BULLY6_{ij} = \nu_0^{BULLY6} + \lambda_1^{BULLY6} Being\_Bullied_{ij} + \varepsilon_{ij}^{BULLY6}, \\
SEFF1_{ij} = \nu_0^{SEFF1} + \lambda_1^{SEFF1} Self\_Efficacy_{ij} + \varepsilon_{ij}^{SEFF1}, \\
SEFF2_{ij} = \nu_0^{SEFF2} + \lambda_1^{SEFF2} Self\_Efficacy_{ij} + \varepsilon_{ij}^{SEFF2}, \\
SEFF3_{ij} = \nu_0^{SEFF3} + \lambda_1^{SEFF3} Self\_Efficacy_{ij} + \varepsilon_{ij}^{SEFF3}, \\
SEFF4_{ij} = \nu_0^{SEFF4} + \lambda_1^{SEFF4} Self\_Efficacy_{ij} + \varepsilon_{ij}^{SEFF4}, \\
MATH_{Wij} = \nu_{0j}^{MATH} + \lambda_{W1}^{MATH} Academic\_Achievement_{Wij} + \varepsilon_{ij}^{MATH}, \\
SCIENCE_{Wij} = \nu_{0j}^{SCIENCE} + \lambda_{W1}^{SCIENCE} Academic\_Achievement_{Wij} + \varepsilon_{ij}^{SCIENCE}, \\
READ_{Wij} = \nu_{0j}^{READ} + \lambda_{W1}^{READ} Academic\_Achievement_{Wij} + \varepsilon_{ij}^{READ}.
\end{cases}
\tag{20}
$$

Level 1 (student) structural model:

$$
\begin{cases}
Academic\_Achievement_{Wij} = \mu_{0j}^{Acad\_Achiev} + \beta_{W1}^{Acad\_Achiev} Parents\_Support_{ij} + \\
\qquad \beta_{W2}^{Acad\_Achiev} Being\_Bullied_{ij} + \beta_{W3}^{Acad\_Achiev} Self\_Efficacy_{ij} + \beta_{W4}^{Acad\_Achiev} SES_{ij} + \\
\qquad \beta_{W5}^{Acad\_Achiev} Gender_{ij} + \beta_{W6}^{Acad\_Achiev} Language\_at\_Home_{ij} + \zeta_{ij}^{Acad\_Achiev}, \\
Self\_Efficacy_{ij} = \mu_{0j}^{Self\_Eff} + \beta_{W1}^{Self\_Eff} Parents\_Support_{ij} + \\
\qquad \beta_{W2}^{Self\_Eff} Being\_Bullied_{ij} + \zeta_{ij}^{Self\_Eff}.
\end{cases}
\tag{21}
$$

The student-level SES variable is specified as:

$$
SES_{Wij} = SES_{Bj} + \varepsilon_{ij}^{SES}. \tag{22}
$$

Level 2 (school) measurement model:

$$\begin{cases} MATH_{Bj} = \nu_{0j}^{MATH} = \nu_0^{MATH} + \lambda_{B1}^{MATH} Academic\_Achievement_{Bj} + \varepsilon_j^{MATH}, \\ SCIENCE_{Bj} = \nu_{0j}^{SCIENCE} = \nu_0^{SCIENCE} + \lambda_{B1}^{SCIENCE} Academic\_Achievement_{Bj} + \varepsilon_j^{SCIENCE}, \\ READ_{Bj} = \nu_{0j}^{READ} = \nu_0^{READ} + \lambda_{B1}^{READ} Academic\_Achievement_{Bj} + \varepsilon_j^{READ}. \end{cases} \tag{23}$$

Level 2 (school) structural model:

$$\begin{aligned} Academic\_Achievement_{Bj} = \mu_0^{Acad\_Achiev} + \beta_{B1}^{Acad\_Achiev} School\_Size_j + \\ \beta_{B2}^{Acad\_Achiev} School\_Type_j + \beta_{B3}^{Acad\_Achiev} Student - Teacher\_Ratio_j + \\ \beta_{B4}^{Acad\_Achiev} SES_j + \zeta_j^{Acad\_Achiev}, \end{aligned} \tag{24}$$

The school-level SES variable is specified as follows:

$$SES_{Bj} = \nu_0^{SES} + \varepsilon_j^{SES}. \tag{25}$$

6 Results

At the time of writing this thesis, the multilevel capabilities of a popular *lavaan* package used for SEM in *R* are limited. Only two-level SEM with random intercepts with all continuous data are allowed, meaning that weights at any level, alternative estimators and categorical variables are not supported. Therefore, the MSEM model was estimated using *Mplus* version 7, integrated into *R* with the *MplusAutomation* package. Scaled student and unscaled school level weights were used in the estimation process. To ensure model identification, plausible values were divided by a constant (30) to keep variances of the variables between 1 and 10, as instructed in *Mplus* User Guide [30]. Since most of the variables were categorical and non-normally distributed, a diagonally weighted least square (DWLS) estimator was employed.

The analysis was performed across 10 datasets, each containing a different set of plausible values (e.g., PV1MATH, PV1SCIE, PV1READ for the first dataset, PV2MATH, PV2SCIE, PV2READ for the second dataset, and so forth). The results were then combined to provide a comprehensive outcome. The model fit indices, presented in Table 3, showed that the model fit the data well. The parameter estimates for the student and school levels are shown in Table 4, graphical model representation is presented in Figure 4. Results of the measurement models are included in Appendix A.

	RMSEA	CFI	TLI	$SRMR_W$	$SRMR_B$
Value	0.048	0.977	0.973	0.042	0.021

Table 3: MSEM model fit indices.

At the student level, the findings demonstrated a significant positive relationship between parental support and students' academic achievement ($\beta = 0.124$, $p < 0.05$), indicating that parents' support plays a major role in helping students succeed in school and enhancing their academic performance. Self-efficacy also had a significant positive impact on students' academic achievement ($\beta = 0.164$,

Construct	Estimate	S.E.	p-value
Student Level			
Direct Effect			
Parents' Support → Academic Achievement	0.124	0.022	0.006
Experience Being Bullied → Academic Achievement	-0.226	0.021	0.000
Self-efficacy → Academic Achievement	0.164	0.024	0.008
SES → Academic Achievement	0.197	0.015	0.000
Gender → Academic Achievement	0.018	0.012	0.193
Language Spoken at Home → Academic Achievement	-0.103	0.015	0.000
Indirect Effect			
Parents' Support → Self-efficacy → Academic Achievement	0.062	0.011	0.008
Experience Being Bullied → Self-efficacy → Academic Achievement	-0.013	0.003	0.025
School Level			
Direct Effect			
School Type → Academic Achievement	-0.308	0.055	0.000
School Size → Academic Achievement	0.193	0.020	0.039
Student-Teacher Ratio → Academic Achievement	-0.132	0.064	0.000
SES → Academic Achievement	0.651	0.045	0.000

Table 4: Parameter estimates of the MSEM model (standardized values).

$p < 0.05$), suggesting that students' belief in themselves and their actions can lead them to achieve better academic results. Moreover, a significant positive relationship was observed between SES and academic achievement ($\beta = 0.197$, $p < 0.05$), indicating that students perform better academically if they come from a higher SES background. Academic achievement was not significantly affected by gender ($\beta = 0.018$, $p > 0.05$).

On the contrary, a significant negative relationship between bullying and academic achievement was observed ($\beta = -0.226$, $p < 0.05$), signifying that students who experience frequent bullying have lower academic achievements. Also, language spoken at home resulted in a negative effect on student achievement ($\beta = -0.103$, $p < 0.05$), meaning that students who speak a language other than the test language perform worse at school. This, in turn, highlights the difficulties students face brought on by language differences and communication barriers.

The mediating role of self-efficacy revealed the statistically significant and positive indirect effect of parents' support on student achievement ($\beta = 0.062$, $p < 0.05$). When parents encourage and support their children, it can increase their belief in their abilities, leading to better academic performance. On the other hand, a significant negative indirect effect of bullying on student achievement ($\beta = -0.013$, $p < 0.05$) was observed. When students experience bullying, it can negatively impact their confidence and belief in themselves, resulting in poorer academic performance.

At the school level, a significant positive relationship between school size and academic achievement was noted ($\beta = 0.193$, $p < 0.05$), suggesting that students perform better academically in larger schools. Schools with a higher number of students, usually located in larger cities, might provide more possibilities and a wider range of resources, which would improve student performance. Additionally, SES demonstrated a positive effect on academic achievement ($\beta = 0.651$, $p < 0.05$), meaning that a higher SES position within the school community is linked to better student results. Conversely, school type showed a significant negative effect on student achievement ($\beta = -0.308$, $p < 0.05$), indicating that

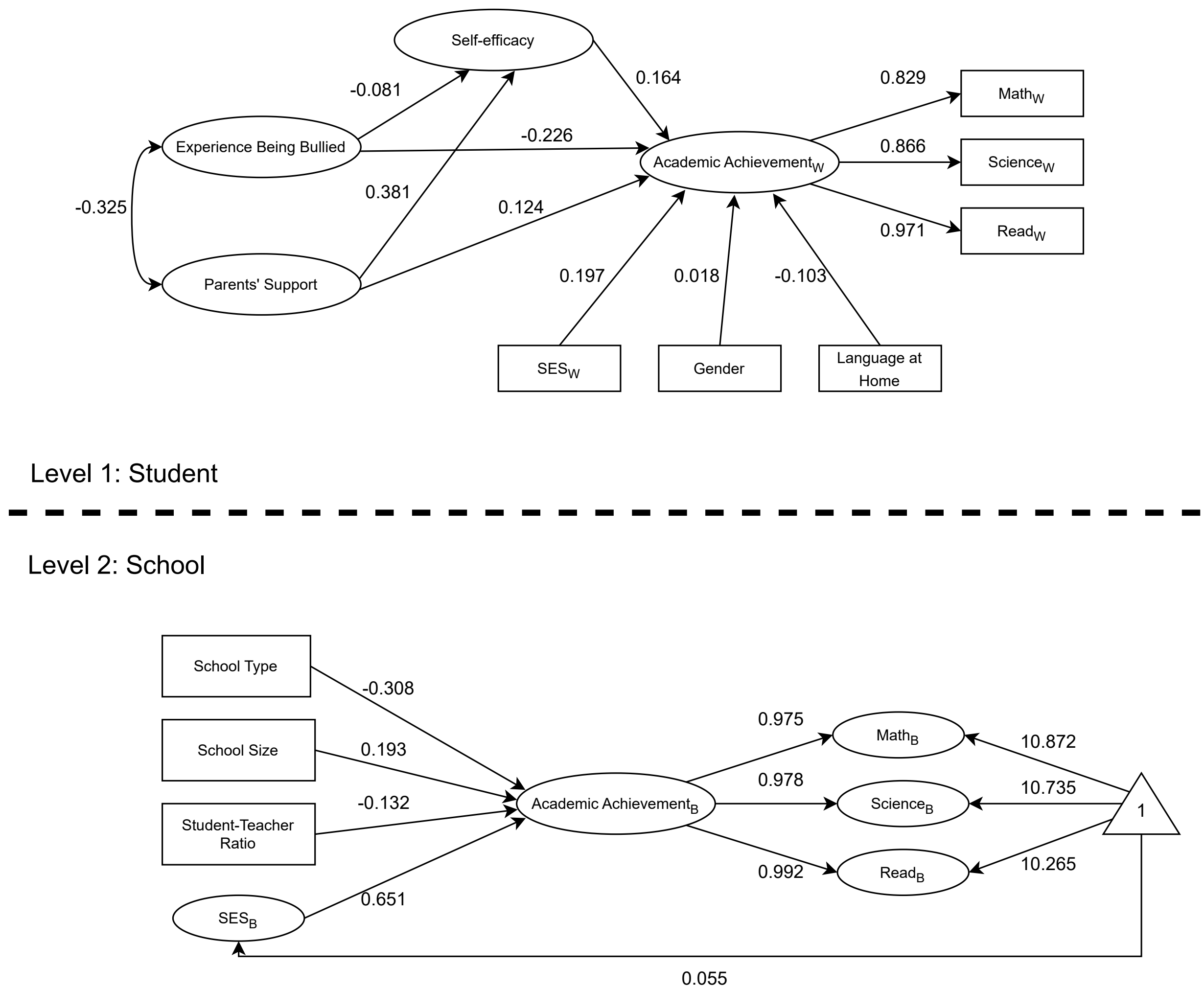

Figure 4: Results of the MSEM model (standardized values).

students from public schools on average achieve lower academic results, compared to private schools. In addition, a significant negative relationship between student-teacher ratio and academic achievement was observed ($\beta = -0.132$, $p < 0.05$), implying that a higher student-teacher ratio is linked to lower average student accomplishments. When the number of teachers is small, it typically results in larger classes, which limits individualized teacher attention and strains student-teacher relations.

7 Alternative tested models

Motivated by insights gained from the literature review, where variables were mentioned to be implemented on both student and school levels, an alternative MSEM model was explored attempting to incorporate latent variables of parental support, experience being bullied, and self-efficacy at both levels via latent decomposition. Although the theoretical framework for this model was supported, there were difficulties implementing it practically using *Mplus* software with DWLS estimator. Due to a large number of categorical variables and weights at both levels, the model could not be identified. To address the issue, an additional model was tested using the Bayes estimator, which is also suggested for estimation with categorical data [15]. However, it was discovered that the this estimator in *Mplus* does not support the use of weights.

8 Conclusions

This thesis utilized multilevel structural equation modeling to investigate the impact of various factors on student achievement in Lithuania, using data from PISA 2018. The study found that students who received higher levels of parental support and came from better socioeconomic backgrounds achieved higher academic results. On the other hand, students who were bullied or spoke a language other than Lithuanian at home tended to perform worse academically. Although gender differences were observed, they were not statistically significant. Self-efficacy, which was identified as a mediator, was found to have a positive impact on student achievements and provided a way to understand the influence of parental support and experiences of bullying on academic achievements further. Therefore, at the student level, parental support, experiences of bullying, self-efficacy, language spoken at home, and socioeconomic status were significant determinants of academic outcomes.

At the school level, larger and private schools, as well as schools with higher socioeconomic status, were associated with better academic outcomes. The student-teacher ratio also played a significant role, with smaller ratios leading to improved academic performance. These findings emphasized the importance of considering both student and school factors to understand student academic success.

The exploration of an alternative model that considered the impact of parental support, bullying, and self-efficacy on both student and school levels presented challenges during the model's application. Although the theoretical framework of the model was promising and was supported by the literature review, it was difficult to implement in practice. This limitation creates opportunities for further research. An alternative approach to address the challenge of incorporating many latent variables at both levels could be to explore simpler models including one of the variables on both levels. Also, as the PISA provides data on students, teachers, and schools, future studies could implement a three-level MSEM model to explore interactions between students, teachers, and schools, providing a more comprehensive understanding of the factors affecting students' academic achievement.

References

[1] Juan Aparicio, Jose M. Cordero, and Lidia Ortiz. "Efficiency Analysis with Educational Data: How to Deal with Plausible Values from International Large-Scale Assessments". In: *Mathematics* 9.13 (2021). ISSN: 2227-7390. DOI: 10.3390/math9131579. URL: https://www.mdpi.com/2227-7390/9/13/1579.

[2] Jolita Buzaitytė-Kašalynienė. "Socialinis palaikymas ir sėkmė mokykloje". In: *Acta Paedagogica Vilnensia* 14 (2005), pp. 145–156.

[3] Diego Carrasco et al. "Indirect effects of bullying on school mathematics achievement in Chile". In: *Studies In Educational Evaluation* 74 (Sept. 2022). DOI: 10.1016/j.stueduc.2022.101172.

[4] Oliver Christ et al. "Advanced Multilevel Modeling for a Science of Groups: A Short Primer on Multilevel Structural Equation Modeling". In: *Group Dynamics Theory Research and Practice* 21 (July 2017), pp. 121–134. DOI: 10.1037/gdn0000065.

[5] Ramazan Atasoy, Ömür Çoban, and Murat Yatağan. "Effect of ICT Use, Parental Support and Student Hindering on Science Achievement: Evidence from PISA 2018". In: *Journal of Learning and Teaching in Digital Age* 7.2 (2022), pp. 127–140. ISSN: 2458-8350. DOI: 10.53850/joltida.945869.

[6] Ömür Çoban. "Relationships between Students' Socioeconomic Status, Parental Support, Students' Hindering, Teachers' Hindering and Students' Literacy Scores: PISA 2018". In: *World Journal of Education* 10 (Aug. 2020), p. 45. DOI: 10.5430/wje.v10n4p45.

[7] Gražina Čiuladienė and Antanas Valantinas. "Tėvų įsitraukimo į mokyklos gyvenimą, mokymosi pasiekimų ir mokytojo darbo vertinimo sąsajos". In: *Socialinis darbas* 2 (2013), pp. 308–318. DOI: 10.13165/SD-13-12-2-09.

[8] Christine DiStefano and Tiejun Zhang. "A Primer for Using Multilevel Confirmatory Factor Analysis Models in Educational Research". In: Jan. 2022, pp. 11–28. ISBN: 978-981-16-9141-6. DOI: 10.1007/978-981-16-9142-3_2.

[9] Thurston Domina. "Leveling the Home Advantage: Assessing the Effectiveness of Parental Involvement in Elementary School". In: *Sociology of Education* 78.3 (2005), pp. 233–249. DOI: 10.1177/003804070507800303. URL: https://doi.org/10.1177/003804070507800303.

[10] Rasa Erentaitė et al. "Profiles of Achievement Motivation and Performance in Middle School: Links to Student Background and Perceived Classroom Climate". In: *Frontiers in Psychology* 13 (May 2022). DOI: 10.3389/fpsyg.2022.820247.

[11] John Freeman et al. "The relationship of schools to emotional health and bullying". In: *International journal of public health* 54 Suppl 2 (Sept. 2009), pp. 251–9. DOI: 10.1007/s00038-009-5421-9.

[12] Alfred Sakyi Fu Chen and Ying Cui. "Linking student, home, and school factors to reading achievement: the mediating role of reading self-efficacy". In: *Educational Psychology* 41.10 (2021), pp. 1260–1279. DOI: 10.1080/01443410.2021.1953445. URL: https://doi.org/10.1080/01443410.2021.1953445.

[13] Su Gao et al. "The mediation effect of student self-efficacy between teaching approaches and science achievement: findings from 2011 TIMSS US data". In: *Social Psychology of Education* 23 (Apr. 2020), pp. 385–410. DOI: 10.1007/s11218-019-09534-1.

[14] Andrew Gelman. "Multilevel (Hierarchical) Modeling: What It Can and Cannot Do". In: *Technometrics* 48.3 (2006), pp. 432–435. DOI: 10.1198/004017005000000661.

[15] Jana Holtmann et al. "A Comparison of ML, WLSMV, and Bayesian Methods for Multilevel Structural Equation Models in Small Samples: A Simulation Study". In: *Multivariate Behavioral Research* 51 (Sept. 2016), pp. 1–20. DOI: 10.1080/00273171.2016.1208074.

[16] J. J Hox. *Multilevel regression and multilevel structural equation modeling*. The Oxford handbook of quantitative methods: Statistical analysis. Oxford University Press, 2013, pp. 281–294.

[17] Joop Hox, Cora Maas, and Matthieu Brinkhuis. "The effect of estimation method and sample size in multilevel structural equation modeling". In: *Statistica Neerlandica* 64 (May 2010), pp. 157–70. DOI: 10.1111/j.1467-9574.2009.00445.x.

[18] Hsien-Yuan Hsu et al. "Detecting Misspecified Multilevel Structural Equation Models with Common Fit Indices: A Monte Carlo Study". In: *Multivariate Behavioral Research* 50 (Mar. 2015), pp. 197–215. DOI: 10.1080/00273171.2014.977429.

[19] Aukščiausioji audito institucija. "Ar pokyčiai švietime lemia geresnius mokinių pasiekimus". In: *Valstybinio audito ataskaita* (Sept. 2020).

[20] Violeta Jegelevičienė, Odeta Merfeldaitė, and Asta Railienė. "Mokinių mokymosi motyvacijos silpnėjimo priežastys: tėvų požiūris". In: *Pedagogika* 124 (2016), pp. 117–128. DOI: http://dx.doi.org/10.15823/p.2016.56.

[21] Karl Jöreskog. "Structural Equation Modeling with Ordinal Variables using LISREL". In: *IMS Lecture Notes-Monograph Series* 24 (Jan. 1994). DOI: 10.1214/lnms/1215463803.

[22] Chiaki Konishi et al. "Do School Bullying and Student-Teacher Relations Matter for Academic Achievement?: A Multilevel Analysis". In: *Canadian Journal of School Psychology* 25 (Mar. 2010), pp. 19–39. DOI: 10.1177/0829573509357550.

[23] Inga Laukaityte and Marie Wiberg. "Using plausible values in secondary analysis in large–scale assessments". In: *Communications in Statistics - Theory and Methods* 46 (Dec. 2016). DOI: 10.1080/03610926.2016.1267764.

[24] Cheng-Hsien Li. "Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares". In: *Behavior research methods* 48 (July 2015), pp. 936–949. DOI: 10.3758/s13428-015-0619-7.

[25] Julia Mang et al. "Sampling weights in multilevel modelling: an investigation using PISA sampling structures". In: *Large-scale Assessments in Education* 9 (Mar. 2021). DOI: 10.1186/s40536-021-00099-0.

[26] G.A. Marcoulides and R.E. Schumacker. "Multilevel Modeling with SEM". In: *New Developments and Techniques in Structural Equation Modeling*. Taylor & Francis, 2001. Chap. 4, pp. 89–127. ISBN: 9781135657819.

[27] Bart Meuleman. "Multilevel Structural Equation Modeling for Cross-National Comparative Research". In: *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 71 (June 2019), pp. 129–155. DOI: 10.1007/s11577-019-00605-x.

[28] Brigita Mieziene et al. "Mokymosi pasiekimus sąlygojantys ir psichosocialines rizikas mažinantys mokyklos bendruomenės ir lyderystės veiksniai: pagrindiniai projekto rezultatai ir rekomendacijos". (Nov. 2021).

[29] Diana Mindrila. "Maximum Likelihood (ML) and Diagonally Weighted Least Squares (DWLS) Estimation Procedures: A Comparison of Estimation Bias with Ordinal and Multivariate Non-Normal Data". In: *International Journal of Digital Society* 1 (Mar. 2010), pp. 60–66. DOI: 10.20533/ijds.2040.2570.2010.0010.

[30] L.K. Muthén and B.O. Muthén. *Mplus User's Guide. 8th edition*. Apr. 2017.

[31] OECD. "Lithuania". In: *PISA 2018 Results (Volume I), OECD Publishing, Paris* (2019).

[32] OECD. *PISA 2018 Assessment and Analytical Framework*. OECD Publishing, Paris, 2019. DOI: https://doi.org/10.1787/b25efab8-en.

[33] OECD. "Sample Design". In: *PISA 2018 Technical Report Chapter 4, OECD Publishing, Paris* (2020).

[34] OECD. "Scaling PISA data". In: *PISA 2018 Technical Report Chapter 9, OECD Publishing, Paris* (2020).

[35] OECD. "Survey Weighting and the Calculation of Sampling Variance". In: *PISA 2018 Technical Report Chapter 8, OECD Publishing, Paris* (2020).

[36] Sophia Rabe-Hesketh, Anders Skrondal, and Xiaohui Zheng. "Multilevel Structural Equation Modeling". In: *Handbook of Latent Variable and Related Models* (Dec. 2007). DOI: 10.1016/B978-044452044-9/50013-6.

[37] Mijke Rhemtulla, Patricia Brosseau-Liard, and Victoria Savalei. "When Can Categorical Variables Be Treated as Continuous? A Comparison of Robust Continuous and Categorical SEM Estimation Methods Under Suboptimal Conditions". In: *Psychological methods* 17 (July 2012), pp. 354–73. DOI: 10.1037/a0029315.

[38] B.C. Silva, C.M. Bosancianu, and L. Littvay. *Multilevel Structural Equation Modeling*. Quantitative Applications in the Social Sciences. SAGE Publications, 2019. ISBN: 9781544323053.

[39] Laura Stapleton. "The Incorporation of Sample Weights Into Multilevel Structural Equation Models". In: *Structural Equation Modeling: A Multidisciplinary Journal* 9 (Nov. 2009), pp. 475–502. DOI: 10.1207/S15328007SEM0904_2.

[40] Daiva Šukytė. "Patyčių problema mokykloje ir prevencija, Švietimo problemos analizė". (Sept. 2014).

[41] Jaclyn Broadbent Toni Honicke and Matthew Fuller-Tyszkiewicz. "The self-efficacy and academic performance reciprocal relationship: the influence of task difficulty and baseline achievement on learner trajectory". In: *Higher Education Research & Development* 0.0 (2023), pp. 1–18. DOI: 10.1080/07294360.2023.2197194. URL: https://doi.org/10.1080/07294360.2023.2197194.

[42] Brad Verhulst and Michael Neale. “Best Practices for Binary and Ordinal Data Analyses”. In: *Behavior Genetics* 51 (May 2021), pp. 204–214. DOI: 10.1007/s10519-020-10031-x.

[43] Chuang Wang and Ting Sun. “Relationship between self-efficacy and language proficiency: A meta-analysis”. In: *System* 95 (Dec. 2020), p. 102366. DOI: 10.1016/j.system.2020.102366.

[44] Christopher Winship and Robert D. Mare. “Regression Models with Ordinal Variables”. In: *American Sociological Review* 49.4 (1984), pp. 512–525.

[45] Algirdas Zabulionis. “Tarptautinio švietimo tyrimo OECD PISA Lietuvos ir kaimyninių šalių duomenų tikslinė antrinė analizė”. (2020).

Appendix A

Latent variable	Indicator	Factor loading
Student level		
Parents Support		
	PS1	0.864*
	PS2	0.889*
	PS3	0.914*
Experience Being Bullied		
	BULLY1	0.821*
	BULLY2	0.860*
	BULLY3	0.945*
	BULLY4	0.935*
	BULLY5	0.942*
	BULLY6	0.885*
Self-efficacy		
	SEFF1	0.743*
	SEFF2	0.619*
	SEFF3	0.707*
	SEFF4	0.795*
Academic Achievement		
	MATH_W	0.829*
	SCIENCE_W	0.866*
	READ_W	0.971*
School level		
Academic Achievement		
	MATH_B	0.975*
	SCIENCE_B	0.978*
	READ_B	0.992*

*Note: * p-value<0.05.*

Results of the measurement models (standardized values).

Appendix B

```
library(dplyr)
library(haven)
library(intsvy)
library(misty)
library(MplusAutomation)

# Reading student data
data_students <- read_sav("CY07_MSU_STU_QQQ.sav")
# Reading school data
data_school <- read_sav("CY07_MSU_SCH_QQQ.sav")

# Filtering and selecting data
data_students1 <- data_students %>%
                  filter(CNT=="LTU") %>%
                  select(CNTSCHID, ST123Q02NA, ST123Q03NA, ST123Q04NA, ST038Q03NA,
                  ST038Q04NA, ST038Q05NA, ST038Q06NA, ST038Q07NA, ST038Q08NA, ESCS,
                  ST188Q01HA, ST188Q03HA, ST188Q06HA, ST188Q07HA, ST022Q01TA,
                  PV1MATH, PV2MATH, PV3MATH, PV4MATH, PV5MATH, PV6MATH, PV7MATH,
                  PV8MATH, PV9MATH, PV10MATH, PV1READ, PV2READ, PV3READ, PV4READ,
                  PV5READ, PV6READ, PV7READ, PV8READ, PV9READ, PV10READ, PV1SCIE,
                  PV2SCIE, PV3SCIE, PV4SCIE, PV5SCIE, PV6SCIE, PV7SCIE, PV8SCIE,
                  PV9SCIE, PV10SCIE, ST004D01T, W_FSTUWT)

# Converting gender and language to dummy variables
data_students1[, "ST022Q01TA"] = data_students1[, "ST022Q01TA"] - 1
data_students1[, "ST004D01T"] = 2 - data_students1[, "ST004D01T"]

data_school1 <- data_school %>%
                filter(CNT=="LTU") %>%
                select(CNTSCHID, SCHLTYPE, STRATIO, SCHSIZE, W_SCHGRNRABWT) %>%
                mutate(SCHLTYPE = ifelse(SCHLTYPE==3,1,0))

full_data <- merge(data_students1, data_school1, by="CNTSCHID")

# Exploratory data analysis
q <- quantile(full_data$ESCS, c(.20, .40, .60, .80), na.rm = T)
```

```
full_data1 <- full_data %>% mutate(CNT=1,
                            SES = case_when(
                            ESCS < q[[1]] ~ 0,
                            ESCS >=q[[2]] & ESCS < q[[3]] ~ 1,
                            ESCS >= q[[4]] ~ 2))

# Weighted mean calculation
# Sample
pisa.mean.pv(pvlabel = paste0("PV",1:10,"MATH"), by = "CNT", data = full_data1)
pisa.mean.pv(pvlabel = paste0("PV",1:10,"SCIE"), by = "CNT", data = full_data1)
pisa.mean.pv(pvlabel = paste0("PV",1:10,"READ"), by = "CNT", data = full_data1)

# By gender
pisa.mean.pv(pvlabel = paste0("PV",1:10,"MATH"), by = "ST004D01T", data = full_data1)
pisa.mean.pv(pvlabel = paste0("PV",1:10,"READ"), by = "ST004D01T", data = full_data1)
pisa.mean.pv(pvlabel = paste0("PV",1:10,"SCIE"), by = "ST004D01T", data = full_data1)

# By school type
pisa.mean.pv(pvlabel = paste0("PV",1:10,"MATH"), by = "SCHLTYPE", data = full_data1)
pisa.mean.pv(pvlabel = paste0("PV",1:10,"READ"), by = "SCHLTYPE", data = full_data1)
pisa.mean.pv(pvlabel = paste0("PV",1:10,"SCIE"), by = "SCHLTYPE", data = full_data1)

# By language spoken at home
pisa.mean.pv(pvlabel = paste0("PV",1:10,"MATH"), by = "ST022Q01TA", data = full_data1)
pisa.mean.pv(pvlabel = paste0("PV",1:10,"SCIE"), by = "ST022Q01TA", data = full_data1)
pisa.mean.pv(pvlabel = paste0("PV",1:10,"READ"), by = "ST022Q01TA", data = full_data1)

# By SES group
pisa.mean.pv(pvlabel = paste0("PV",1:10,"MATH"), by = "ESCS_P", data = full_data1)
pisa.mean.pv(pvlabel = paste0("PV",1:10,"SCIE"), by = "ESCS_P", data = full_data1)
pisa.mean.pv(pvlabel = paste0("PV",1:10,"READ"), by = "ESCS_P", data = full_data1)


pattern_math <- "PV\\d+[MATH]"
pattern_scie <- "PV\\d+[SCIE]"
pattern_read <- "PV\\d+[READ]"

columns_math <- grep(pattern_math, names(full_data), value=TRUE)
columns_scie <- grep(pattern_scie, names(full_data), value=TRUE)
columns_read <- grep(pattern_read, names(full_data), value=TRUE)
```

```
# ICC ranges
min(multilevel.icc(full_data[,columns_math], cluster = full_data$CNTSCHID))
max(multilevel.icc(full_data[,columns_math], cluster = full_data$CNTSCHID))

min(multilevel.icc(full_data[,columns_read], cluster = full_data$CNTSCHID))
max(multilevel.icc(full_data[,columns_read], cluster = full_data$CNTSCHID))

min(multilevel.icc(full_data[,columns_scie], cluster = full_data$CNTSCHID))
max(multilevel.icc(full_data[,columns_scie], cluster = full_data$CNTSCHID))

# Performing listwise deletion and selecting clusters with at least 10 students
full_data2 <- full_data[complete.cases(full_data ), ] %>%
              group_by(CNTSCHID) %>%
              filter(n() > 9)

# Testing data normality
normality <- apply(full_data2, 2, shapiro.test)
normality

# Creating 10 datasets to estimate the model 10 times
list_of_datasets <- list()

for (i in 1:10) {
math_column <- grep(paste0("PV", i, "MATH"), columns_math, value=TRUE)
scie_column <- grep(paste0("PV", i, "SCIE"), columns_scie, value=TRUE)
read_column <- grep(paste0("PV", i, "READ"), columns_read, value=TRUE)

split_dataset <- full_data2[, c(setdiff(names(full_data2),
                              c(columns_math, columns_scie, columns_read)),
                              math_column, scie_column, read_column)]

list_of_datasets[[paste("dataset_", i, sep="")]] <- split_dataset
}



# Creating MSEM model
MSEM_model <- mplusObject(
  TITLE = "STUDENT ACHIEVEMENT MSEM MODEL",
```

```
  ! Naming variables
VARIABLE = "NAMES = CNTSCHID PS1 PS2 PS3
                    BULLY1 BULLY2 BULLY3 BULLY4 BULLY5 BULLY6
                    SES SEFF1 SEFF2 SEFF3 SEFF4
                    LANG GENDER WSTU
                    SCHLTYPE STRATIO SCHSIZE WSCH
                    MATH SCIENCE READ;

  ! Specifying categorical variables
             CATEGORICAL = PS1 PS2 PS3
                           SEFF1 SEFF2 SEFF3 SEFF4
                           BULLY1 BULLY2 BULLY3 BULLY4 BULLY5 BULLY6;

  ! Specifying student level only variables
             WITHIN = PS1 PS2 PS3
                      BULLY1 BULLY2 BULLY3 BULLY4 BULLY5 BULLY6
                      SEFF1 SEFF2 SEFF3 SEFF4
                      GENDER LANG;

  ! Specifying school level only variables
             BETWEEN = SCHLTYPE STRATIO SCHSIZE;

  ! Clustering by school
             CLUSTER IS CNTSCHID;

  ! Denoting scaled student and unscaled school level weights
             WEIGHT = WSTU;
             WTSCALE = ECLUSTER;

             BWEIGHT = WSCH;
             BWTSCALE = UNSCALED",

  ! Dividing subject variables by a constant
DEFINE = "MATH = MATH/30;
          READ = READ/30;
          SCIENCE = SCIENCE/30;

  ! Selecting an estimator for categorical data
ANALYSIS = "TYPE IS TWOLEVEL;
            ESTIMATOR=WLSMV;",
```

```
  ! Specifying model for student (within) and school (between) levels
MODEL = "%WITHIN%

  ! Student level measurement model
      PAR_SUPP BY PS1 PS2 PS3;
      BULLYING BY BULLY1 BULLY2 BULLY3 BULLY4 BULLY5 BULLY6;
      SELF_EFF BY SEFF1 SEFF3 SEFF4 SEFF5;
      ACADEM_W BY MATH SCIENCE READ;

  ! Student level structural model
      ACADEM_W ON PAR_SUPP
                  BULLYING
                  SELF_EFF (a)
                  SES
                  GENDER
                  LANG;

      SELF_EFF ON PAR_SUPP (b)
                  BULLYING (c);

      PAR_SUPP WITH BULLYING;

         %BETWEEN%
      SES;

  ! School level measurement model
      ACADEM_B BY MATH SCIENCE READ;

  ! School level structural model
      ACADEM_B ON SES
                  SCHLTYPE
                  STRATIO
                  SCHSIZE;",

  ! Specifying indirect effects
MODELCONSTRAINT = "NEW (ab ac);
      ab = a*b;
      ac = a*c;",

  ! Requesting standardized parameters
OUTPUT = "STDYX",
```

```
  rdata = list_of_datasets,
  imputed = TRUE) # performing analysis on 10 datasets at the same time

# Estimating the MSEM model
model_fit <- mplusModeler(MSEM_model, modelout = "MSEM_model.inp",
                          run = TRUE, quiet = FALSE)

# Extracting aggregated standardized model parameters
model_fit$results$parameters$stdyx.standardized

# Extracting aggregated model fit indices
model_fit$results$summaries


# Model that could not be identified
MSEM_trial <- mplusObject(
  TITLE = "STUDENT ACHIEVEMENT MSEM MODEL TRIAL",

  VARIABLE = "NAMES = CNTSCHID PS1 PS2 PS3
                      BULLY1 BULLY2 BULLY3 BULLY4 BULLY5 BULLY6
                      SES SEFF1 SEFF2 SEFF3 SEFF4
                      LANG GENDER WSTU
                      SCHLTYPE STRATIO SCHSIZE WSCH
                      MATH SCIENCE READ;

               CATEGORICAL = PS1 PS2 PS3
                             SEFF1 SEFF2 SEFF3 SEFF4
                             BULLY1 BULLY2 BULLY3 BULLY4 BULLY5 BULLY6;

               WITHIN = GENDER LANG;

               BETWEEN = SCHLTYPE STRATIO SCHSIZE;

               CLUSTER IS CNTSCHID;

               WEIGHT = WSTU;
               WTSCALE = ECLUSTER;

               BWEIGHT = WSCH;
               BWTSCALE = UNSCALED",
```

```
DEFINE = "MATH = MATH/30;
          READ = READ/30;
          SCIENCE = SCIENCE/30;

ANALYSIS = "TYPE IS TWOLEVEL;
            ESTIMATOR=WLSMV; !BAYES;",

MODEL = "%WITHIN%
      PAR_SUP_W BY PS1 PS2 PS3;
      BULLY_W BY BULLY1 BULLY2 BULLY3 BULLY4 BULLY5 BULLY6;
      SELF_EFF_W BY SEFF1 SEFF3 SEFF4 SEFF5;
      ACADEM_W BY MATH SCIENCE READ;

      ACADEM_W ON PAR_SUP_W
                  BULLY_W
                  SELF_EFF_W (a)
                  SES
                  GENDER
                  LANG;

      SELF_EFF_W ON PAR_SUP_W (b)
                  BULLY_W (c);

         %BETWEEN%
      SES;
      PAR_SUP_B BY PS1 PS2 PS3;
      BULLY_B BY BULLY1 BULLY2 BULLY3 BULLY4 BULLY5 BULLY6;
      SELF_EFF_B BY SEFF1 SEFF3 SEFF4 SEFF5;
      ACADEM_B BY MATH SCIENCE READ;

      ACADEM_B ON PAR_SUP_B
                  BULLY_B
                  SELF_EFF_B
                  SES
                  SCHLTYPE
                  STRATIO
                  SCHSIZE;",

MODELCONSTRAINT = "NEW (ab ac);
      ab = a*b;
      ac = a*c;",
```

```
OUTPUT = "STDYX",
rdata = list_of_datasets,
imputed = TRUE)
```