VILNIAUS UNIVERSITETAS

MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Magistro baigiamasis darbas

Lietuvos moksleivių pažangumas matematikos srityje: PISA duomenų analizė

The analysis of factors which affect school students knowledge of mathematics in Lithuania: indications from PISA

Daniel Lavskij

VILNIUS 2020

MATEMATIKOS IR INFORMATIKOS FAKULTETAS
STATISTINĖS ANALIZĖS KATEDRA

Darbo vadovas Prof. habil. dr Rimas Norvaiša ________________

Darbas apgintas 2020-01-13

Registravimo NR. ________________

Lietuvos moksleivių pažangumas matematikos srityje: PISA duomenų analizė

Santrauka

Šiame darbe buvo tiriama, ar grynos matematikos poveikis turi didesnę įtaką negu taikomosios matematikos poveikis Lietuvos moksleivių PISA matematikos testų rezultatams. Analizė rodo, kad grynos matematikos indeksas turi teigiamą ir nuoseklų ryšį su Lietuvos moskleivių PISA matematikos testų raštingumo balais. Penki pritaikyti modeliai pagrindžia hipotezę: grynos matematikos poveikio indekso padidėjimas vienu vienetu didina studentų matematikos raštingumą 18–27 balais. Toks ryšys išlaikomas ir atsižvelgiant į moksleivių socialinį-ekonominį statusą bei klasę. Tyrimo hipotezė galioja visiems matematikos poskyriams. Taikomosios matematikos poveikis PISA matematikos testui visais atvejais buvo neigiamas ir dažniausiai nereikšmingas. Taikomosios matematikos poveikis matematikos raštingumui Lietuvoje yra "kvadratinis" ir primena kvadratinį sąryšį: jis daro teigiamą įtaką iki kol pasiekia tam tikrą savo indekso lygį ir vėliau matematinis raštingumas pradėda mažėti. Latvijos ir Estijos analizė rodo, grynosios ir taikomosios matematikos indeksų poveikis yra panašūs į Lietuvos atvejį.

Raktiniai žodžiai : (Matematinis raštingumas, Hierarchinis tiesinis modelis, PISA duomenų analizė, taikomosios matematikos poveikis, Lietuvos moksleivių matematikos žinios, PISA duomenų analizė)

The analysis of factors which affect school students knowledge of mathematics in Lithuania: indications from PISA

Abstract

This study examined whether the exposure to pure mathematics has bigger impact on PISA mathematics literacy scores compared to exposure to applied mathematics in Lithuania. Results suggest that exposure to pure mathematics has a positive and consistent relationship with mathematics literacy score on population level in Lithuania. Five fitted models support the hypothesis: a one unit increase in index of exposure to pure mathematics will increase student mathematics literacy score by 18-27 points. The relationship is resilient even after controlling for students' socioeconomic status and grade. The hypothesis of the study holds for all sub-areas of mathematics. Estimate of exposure to applied mathematics coefficient was negative in all cases and usually insignificant. The negative relationship of exposure to applied mathematics with mathematics literacy score in Lithuania appear to be "semi-quadratic": it has a positive relationship with mathematics literacy score, but after a certain level of exposure of applied mathematics index is reached ML score starts to decrease. The analysis of Latvia and Estonia shows that index of exposure to applied mathematics follows similar pattern.

Key words : (Mathematics literacy, hierarchical linear modeling, exposure to pure mathematics, exposure to applied mathematics, Lithuania students' knowledge of mathematics, survey data analysis, PISA data analysis)

Contents

1 Introduction **2**

2 Background **5**

3 Data, its origins and features **9**

3.1 Sampling design . . . 9

3.2 The assessment method . . . 11

3.3 Survey results . . . 12

3.4 Reported values . . . 13

3.4.1 Plausible values . . . 14

3.4.2 Regressors . . . 15

3.4.3 Levels of analysis . . . 16

4 Methodology **18**

4.1 Specification of various models . . . 18

4.2 Weights . . . 20

4.3 Analysis with plausible values . . . 20

5 Practical part **21**

5.1 Exploratory analysis . . . 21

5.2 Modeling step . . . 23

6 Results **27**

6.1 A closer look at the Lithuania: results from different models . . . 27

6.2 Comparison with other countries . . . 32

7 Conclusions **35**

8 Appendix **36**

1 Introduction

School students from Lithuania tend to have lower results in international PISA mathematics tests compared to OECD average. Starting with 2006, when Lithuanian students scored 486 and gap with OECD average was only 4 points [OECD, 2007], the difference extended to 19 points in 2009 [OECD, 2010]. While 2006 results had shown no statistical difference, 2009 results were significant, meaning that Lithuanian students average was significantly lower that OECD average that year. Distinction between Lithuania results and OECD average has shrunk to 8 points according to 2018 data, but still is significantly lower that OECD average [OECD, 2019]. Different results come from TIMMS data, where Lithuanian student average score in mathematics during 2003-2015 year had grown from 502 to 511 points with 2015 year result being significantly higher than TIMMS 500 points average [Mullis et al., 2016]. These results should be treated more carefully because the list of participants goes beyond Europe and OECD countries. Furthermore, 2012 – 2015 PISA data shows that Lithuania had 25,4 – 26 per cent of all students below baseline proficiency level in mathematics [OECD, 2013b, 2016a]. That mean every fourth student can only answer questions which involve familiar context with all the relevant information given; carry out routine tasks with all instructions present; perform actions that are obvious according to given task [OECD, 2013a].

According to Valdemaras Razumas, Vice Minister of education and sports in Lithuania opinion, one of the main reasons why Lithuanian students perform how they do is high level of social exclusion between students in cities and regions; another reason was stated to be insufficient quality of teaching in the regions [LRT, 2019]. In addition, not only those two reasons matter but also the motivations of students – students from city schools can choose from a wide variety of after class activities, while small-regional educations systems suffer from insufficient funding needed to ensure such possibilities to students.

The results from OECD raised some questions to government and educational system, which was under reform since 90's. The effects of it are child-centered paradigm and prioritization of general skills acquired by students through educational curriculum, while the role of teachers in educational process was criticized [Norvaiša, 2019]. The balance between learning and teaching paradigms could be explored by didactics of mathematics. However, Lithuania lacks experience in this particular domain: there is no local organization that works in this area; poor financial support e only two universities in the country work with didactics in mathematics as a scientific discipline; a lack of incentives to perform scientific researches in all areas of didactics of mathematics, such as teaching quality, students learning possibilities and exogenous factors, such as home and school atmosphere [Narkevičienė and Novikienė, 2018].

Besides educational system's political background, the main questions are still here – what should be done in Lithuania differently so that students would perform better in mathematics on average and how to minimize the percentage of students below baseline proficiency level in mathematics. The common knowledge emphasizes the importance of home and school environment at lower proficiency levels, while attitude towards mathematics and self-confidence comes make greater impact on higher proficiency levels. PISA publications based on 2012 data underline the importance of "universal" factor – exposure to pure mathematics which appears to have a strong relationship with final PISA mathematics literacy score [OECD, 2016b]. A lack of such type of mathematics in Lithuania was stated by some students. Mainly they stress the way mathematics is taught in the schools – with no understanding of what actually lies behind the formulas but only blind application of it [Tiešytė et al., 2001]. Content analysis of mathematical textbooks confirms the presence of a problem with an example how the topic of "powers and integer exponents" was presented in different books from 1982 to 2013: the earliest books presented the definition of a concept, properties and proof of properties with specific numbers, while latest book aims its attention only on computation process of particular topic without explaining the meaning or even agreement on formula [Norvaiša, 2019]. Therefore, the hypothesis of this work is to check whether exposure to pure mathematics has bigger influence on standardized PISA test results in Lithuania compared with exposure to applied mathematics.

A proper analysis of student performance in mathematics using PISA survey data suggest implementation of various data analysis techniques. Considering the features of survey data, the analysis of factors was performed using multilevel models which are able to account for differences that appear due to sampling design. As multilevel models suggest hierarchical structure of data, it is important to choose whether or not to scale the weights. Analysis without scaling can result in biased estimates, while application of scaling techniques can over-correct bias, therefore it is advised to perform sensitivity analysis. In this work sensitivity analysis is performed by comparing results from scaled weights and conditional weights. As a "sanity" check, linear regression model was also fitted in order to compare signs and significance levels of estimates. In addition, PISA survey report results as a combination of plausible values, so this work will also overview the techniques of working with such data.

Results suggest that exposure to pure mathematics has a positive and consistent relationship with mathematics literacy score on population level in Lithuania. Such relationship is statistically significant, while exposure to applied mathematics has negative (quadratic form) relationship which is usually insignificant. It is supported by the multilevel models for survey data with different model specification: random intercept with fixed slopes and both random intercepts and slopes. Results from 4 multilevel models and 1 linear regression model support the hypothesis: a one unit increase in index of exposure to pure mathematics will increase student ML score by 18-27 points. The relationship is resilient even after controlling for students' socioeconomic status and grade. The negative relationship of exposure to applied mathematics with mathematics literacy score in Lithuania appear to be "semi-quadratic": it has a positive relationship with mathematics literacy score, but after a certain level of exposure of applied mathematics index is reached e ML starts to decrease. The analysis of Latvia and Estonia shows that index of exposure to applied mathematics follows same pattern, but even greater. This relationship was also confirmed by literature and PISA reports. The hypothesis of the study holds for all sub-areas of mathematics. The estimate of exposure to pure mathematics coefficient vary between 24 and 29 points, meaning that 1 unit increase in index of exposure to pure mathematics will increase student ML score by 24-29 points. Estimate of exposure to applied mathematics coefficient was negative in all cases and usually insignificant.

This work will consist from background section, where context and literature will be reviewed. The second part will describe PISA survey and its main characteristics. After PISA data origins and features section section of used methodology will follow. It will describe specification of models, weights scaling procedures, theory of plausible values and how to work with them. After that practical part will be presented, which was divided into two subsections: exploratory analysis and modeling step. The first will present brief characteristics of data set, while second will describe the workflow. Results and conclusion will be discussed at the end.

2 Background

International organizations are main source of information in a context of student's performance. Such organizations have reliable resources to gather, clean up and prepare data for analysis and, notably, present the comparable results at international level. There are two such worth to mention organizations e International Association for the Evaluation of Educational Achievement (IEA) and The Organisation for Economic Co-operation and Development (OECD). Trends in International Mathematics and Science Study (TIMMS), a flagship study of IEA, focuses on school student's performance in mathematics and science subjects, evaluates the differences how these subjects are being taught between participating countries, collects empirical information about the curriculum and its implementation. This study mainly targets fourth and eighth grade students. TIMMS tends to evaluate student's abilities in mathematical tasks from school curriculum. TIMMS publications analyze a wide variety of topics, e.g. socioeconomic inequalities and educational outcomes [Broer et al., 2019], importance of soft and cross-subject skills, effects of classroom instructions [Kim, 2018]. Programme for international students' assessment (PISA), developed by Organisation for Economic Co-operation and Development (OECD), measures school students' knowledge and skills in science, reading, mathematics. PISA tries to find out whether students can apply what they have learned in school to real life situations. The study mainly focuses 15-year-old school students in each participating county. PISA provides extended studies for each time stamp when data were collected. Latest available studies give insights about student well-being, policies and practices for successful schools, as well as some in-depth reports, such as equity in education and effective teacher policies. In this work author's primary focus is mathematics topic in PISA survey, rather that TIMMS because:

- PISA focuses on mathematics knowledge implementation in real life situations, while TIMMS examines it via test based on school curriculum. This fundamental feature makes PISA results unique and valuable in today's world;
- there are differences between countries education systems, so target audience defined by age could provide more meaningful and comparable information what students are capable of at particular age, rather than grade;
- regularity, which helps countries to track students progress in key aspects of learning;
- the decision to participate in PISA or not is taken by governments, whose policy interested in the results.

The PISA survey, its test design as well as the relevance of results was seriously challenged by academic world. In 2014 more than 100 academics around the globe called OECD to skip the next PISA

2015 cycle [The Guardian, 2014]. The negative consequences of PISA included the escalation of standardized testing for students, teachers, school officials and an unreasonable "rally" to the "top" of PISA rankings, while it is known that results from such tests tend to be imperfect; the three-year assessment cycle has caused a shift to short-term goals, while educational reforms widely known to be long-term in order to achieve any major results; PISA tests only three subjects, while the range of students abilities is way larger – it includes moral, civic, artistic, physical sides and aspects, thereby narrowing our imagination on what education is and how should be taught [Strauss, 2019]. Another portion of criticism was recently stated by Yong Zhao, a Foundation Distinguished Professor in the School of Education at the University of Kansas. He writes: "*The foundation upon which PISA has built its success, has been seriously challenged. First, there is no evidence to justify, let alone prove, the claim that PISA indeed measures skills that are essential for life in modern economies. Second, the claim is an imposition of a monolithic and West-centric view of societies on the rest of the world. Third, the claim distorts the purpose of education*" [Washington Post, 2019]. However, it was nicely stated by R.Ževlys that PISA has many faces – some of them are attractive, whereas others are not. Therefore, various aspects of PISA survey should be used in professional and skillful way in order to achieve quality and valuable results [Želvys et al., 2016].

There is a strong relationship between educational inequality and attainment of educational qualifications that is connected to social status [Breen and Jonsson, 2005]. Students with higher socioeconomic status (SES) (e. g. family financial status, school resources, country income) usually have more learning opportunities and that helps them to perform better [Baker et al., 2002]. Students with low socioeconomic status lack some of the essential skills needed to perform better. Performance differences appear more frequently on those tasks that require deeper knowledge of symbolic and technical operations as well as on tasks which require model-building skills. As a consequence, students' socioeconomic status and differences between schools' socioeconomic distribution of students create additional sources of variation in mathematical literacy e between schools and within school. Across OECD counties, students' SES and differences between schools' composition explain around 9 percent (on average) of the variation in familiarity with mathematics on country level [OECD, 2016b]. The fact that low-SES students are missing needed level of familiarity with mathematics concepts can explain around 19-20 percent of performance differences in Malaysia – students from schools with higher socioeconomic status outperformed students from schools with low socioeconomic status [Thien, 2016]. Differences in students' socioeconomic status have greater significance on students' performance if particular county has unequal distribution of qualified teachers or clustering of privileged students – such factors usually lowers students results [Chiu and Khoo, 2005]. The relationship between SES and student's performance when student and school socioeconomic status are dis-aggregated suggests that increase in school socioeconomic status increases performance of particular school student regardless of their individual SES [Perry and McConney, 2010]. The gap in educational achievements between high and low SES group is also confirmed and shown by TIMMS publications: e.g. despite of Lithuania economy growth during 1995-2015 period, the difference between high- and low- SES students in TIMMS mathematics test results extended from 73 points in 1995 to 90 points in 2015 [Broer et al., 2019]. Numerous PISA reports confirm the existence of a strong link between socioeconomic status and both opportunities to learn and students performance [OECD, 2007, 2010, 2013a, 2016a, 2019].

The four "Opportunity to Learn" (OTL) content variables were identified by Stevens [1993] – coverage, exposure, emphasis and quality of instructional delivery in order to assure that all students receive quality education and support in different fields of it. In order to assess and evaluate the OTL variables PISA 2012 designed special tasks – some representing more traditional tasks, some strongly connected with critical and analytical thinking. The OTL variables considered to be important concept in international surveys like PISA that has strong relationship with overall student performance[Schmidt and Maier, 2009]. PISA results confirm that exposure to pure mathematics have a strong link with performance in mathematics. This relationship increases as mathematical problems become more difficult, while exposure to applied mathematics has weaker relationship with mathematics literacy. Exposure to pure mathematics and learning such type of mathematics through real-world applications are significantly related to all the PISA mathematics literacy scores across 43 countries [Cogan et al., 2019]. The conclusions about performance and pure mathematics relationship were stable even after controlling for other factors that have direct link with performance (e.g. SES, grade level). The connection of applied mathematics and mathematics literacy was not such consistent and in cases when it was significant relationship had quadratic form – with negative coefficient, meaning that after some infection point "Applied mathematics" index was negatively related to performance. Such results support hypothesis that pure mathematics is highly related to mathematics literacy, just as it was found in [OECD, 2016b].

Comprehensive PISA report "Equations and Inequalities. Making mathematics accessible to all" suggests several important conclusions: first of all, it shows that educational systems have a lot of differences in the degree to student's exposure to mathematics concepts and the way those concepts are presented to students. Secondly, such variation is even bigger within country level rather that between. It can be explained by educational system on country level. For example, Lithuania has decentralized education system, where school governance is distributed between different level authorities – state, municipal and, finally, school [OECD, 2017].

While international educational studies provide vast amount of information to analyze, there is a lack of any analysis of mathematical performance conducted on within-country level of Lithuania, especially with PISA data. The targeted analysis of Lithuania which describes the connection of economic and educational variables was performed by [Mikk, 2006] using TIMMS data. The analysis revealed a strong relationship between mathematical results and variables, previously stated by others authors, e.g. students' motivation, economic development of country, teachers related factors and etc. It was also stated that there is a positive relationship of student's self-confidence and mathematics score in TIMMS test, however the estimate was lower compared to countries with higher scores. Another analyses was carried on university students in order to explore a relationship of different variables and students' performance in university. Article by [Murauskas and Radavičius, 2010] suggests that that students' results of mathematical analysis subject are impacted by professor "factor", so the assesment rules must be unified in a more uniform way. This inference is supported by [Radavičius and Samusenko, 2010] where authors suggest that the student's average of grades in university is not appropriate criteria for comparison of students' performance. The main reason of this inference is being the fact that test difficulty between different test/exam composers (professors), so students are "favoured" to choose university courses which are easier. The students' motivation to choose "easier" subjects in university may be connected with their attitude towards mathematics in school. The recent study of Lithuanian and Latvian students' interest in mathematics suggest that almost half of the students dislikes mathematics and show no effort and motivation when dealing with problems that require effort[Cēdere et al., 2018]. Authors also notice that explored cognitive interest variables explain about a half of total variance, so that another reasons of differences in interest are present. Although such results come from non international organization data, it has a connection with PISA results. According to [OECD, 2016b], another source of variation may come from teachers e students taught by particular teacher may have higher mathematics performance as a consequence of higher self-concept, because their teacher differentiates students by their abilities and encourages them to work in small groups. The report also shows some insights about students' interest in mathematics, anxiety towards mathematics and discusses what could be done to ensure that all students have equal or at least similar opportunities in mathematics learning. At OECD level, 53 per cent of students answered that they are interested in mathematics they face in school, while 38 per cent reported that they enjoy to study it. There is a controversy on the achievement-attitude paradox which suggests that mean of achievements level correlate with mean of attitude level towards particular subject, but the sign is opposite to what is expected – greater results are being achieved by students with more negative feelings about that subject rather than opposing; however, it can't be explained through cultural differences, but the answers may be found in alternative questionnaire designs [Lu and Bolt, 2015].

It is clear that educational systems have a lot of differences in the degree to student's exposure to mathematics concepts and the way those concepts are presented to students. Such variation is even bigger within country level rather that between [OECD, 2017]. It can be explained by educational system on country level[1]. The literature review suggest that analysis of students' school performance is a common topic on international level, however not in Lithuania. Therefore the main hypothesis of this study is stated as:

Exposure to pure mathematics has bigger influence on PISA test results in Lithuania compared with applied mathematics.

The main tasks of the work is:

- To examine descriptive statistics of students' performance in mathematics on country level;

[1] Lithuania has decentralized education system, where school governance is distributed between different level authorities – state, municipal and, finally, school.

- To perform complex PISA survey data analysis using various regression methods on country level;
- To compare differences in parameter estimates between models on country level;
- To collate results with other countries.

The additional task will be to study whether the hypothesis of study holds independently from area (capital, city, town, village), school type (gymnasium, secondary school, basic school, other), student socioeconomic status (socially-economically advantaged, "normal" and disadvantaged students). However, the analysis on different levels than country could provide biased estimates and must be treated more carefully because of the features of sampling design, as it described in 3.1.

3 Data, its origins and features

OECD describes The Programme for International Students Assessment (PISA) as:

"*collaborative effort among OECD member countries to measure how well 15-year-old students approaching the end of compulsory schooling are prepared to meet the challenges of today's knowledge societies*" [OECD, 2014, page 22].

PISA collects information every three years and presents results of knowledge in mathematics, science and reading at three different levels – students, schools and countries. PISA 2012 – fifth PISA survey with primary focus on mathematics, while reading, science and problem-solving topics were minor areas of the survey. The assessment of three major topics is combined with explicit information about each student: his approach to learning, home background, learning environment, self-confidence and motivation. In addition, PISA expands the list of factors which may influence student results by addressing additional questionnaires to school principals and parents. Thereby, study examines how particular home and school factors interact with students results and gives insights on factors that actually make a difference. Starting with 2006, Lithuania participated in PISA survey as partner country. In 2012 it was the first time when Lithuania participated in mathematically-targeted PISA survey.

This section will present general ideas of PISA survey. First subsection will present features of sampling design, second e general idea about assessment method, third e brief explanation of survey outcomes, and fourth e an insights about reported values that are main focus of the study.

3.1 Sampling design

Surveys in education usually are drawn with two-stage stratified sample design. First of all, schools with target audience of 15-years old are chosen from a comprehensive national list of all PISA-eligible schools. Then, a simple random sample is drawn from all available students in that school. Therefore, two-stage sampling procedure will affect weights calculation process, while properties of student sample will be impacted by school selection procedure. Main characteristics of PISA 2012 two-stage stratified sample design consisted of:

- First stage – schools are chosen from school sampling frame with probabilities proportional to a number of 15-years-old students enrolled in the particular school. Consequently, larger schools (with bigger PISA-eligible target audience) have had a bigger probability of being selected than small schools. This method is also called systematic probability proportional to size (PPS) sampling [OECD, 2014].

- Second stage – students within sampled schools are chosen with equal probabilities (simple random sample). Particularly, a complete list of 15-years-old PISA-eligible students was prepared for each selected school. Within the framework of PISA, the number of students from each school that will be chosen in the sample was set to 35 (although it could vary between countries if there is an agreement between government and survey officials). Whenever so-called Target Cluster Size (TCS) was set, a sample of students were selected with equal probabilities if school PISA survey target audience was bigger than TCS; in case whether it was equal or smaller than TCS, all PISA-eligible students on the list were selected.

In order to ensure accuracy and precision, PISA selected the sample students using professional principles of scientific sampling, so that the full target population of PISA-eligible 15-years-old were represented in each participating country. The minimum number of schools and students had to be 150 and 4500 respectively. While those two values may be less than defined minimum (e.g. by choosing less schools, but more students, or by choosing all available target population, if it is less than defined size), the TCS minimum threshold was defined to be at least 20 students per school. This requirement was obligatory in order to fulfill a major objective of PISA analysis – the within- and between- school variance components estimation with adequate accuracy. The minimum requirements were fulfilled in Lithuania – the number of selected schools was equal to 216, while the "unweighted" number of students – 4618.

PISA data set includes the survey weights which facilitate the proper analysis and ensure that inferences made on population level are valid. However, weights are not the same across students' e the selection probabilities of students may vary due to complex two-stage survey design which consists from school and student sampling. Another reasons, why weights vary are: the intention of country or government to over- or under-sample particular sectors of school population for international purposes; in-accurateness of information about school size at the time of school sampling; school and student non-response rate; survey weights trimming effect in order to prevent the influence of a small group of students.

The final weight W_{ij} for student j from school i consists from two base weights and five adjustment factors [OECD, 2014, page 133]:

$$W_{ij} = t_{2ij} f_{1i} f_{2ij} f_{1ij}^{A} t_{1i} w_{2ij} w_{1i} \tag{1}$$

where:

- w_{1i} e school base weight, the inverse of the probability that particular school i will be selected in school sample.

- w_{2ij}e student base weight (within-school), the inverse of the probability that student j within the school i will be selected.

- $t_{2ij} f_{1i} f_{2ij} f_{1ij}^{A} t_{1i}$ e adjustment and trimming factors. For more information see [Chapter 8 of PISA 2012 Technical Report, OECD, 2014, pages 132-141].

For simplicity (although, the values of trimming and adjustment factors are unknown) the formula of W_{ij} will be equal to the product of w_{2ij} and w_{1i}.

In order to assure that population parameter estimates are unbiased PISA data should be weighted[OECD, 2009, page 56]. The sub-populations parameter estimates using weights might be still biased, because the survey is designed to represent a country, meaning that there were no sub-populations (by region, school type, student-socioeconomic status) sample design or it is unknown. Therefore, such sub-populations will not represent itself as a different sample, but as a part of country level design.

3.2 The assessment method

International surveys such a PISA often have a complex assessment design. Such complexity comes from variety of developed items which are included in the final tests in order to provide valid and comparable results. The full PISA 2012 paper-based assessment consisted of 110 cognitive mathematics items, 44 reading and 53 science items, which required respectively 270, 90 and 90 minutes of testing time [OECD, 2014]. However, it is unreasonable and even impossible to test each student with whole set of questions due to:

- Student's results may be affected by fatigue which could start from extended testing time; as a consequence, it can bias the survey results.

- School administrations may refuse to free their students for such a long time and that would produce additional undesirable bias in the results because of the reduction on school participation rate [OECD, 2009].

In order to overcome those limitations, the whole set of items were divided into thirteen item clusters (seven mathematics clusters and three clusters for each reading and science). The final standard booklets for the Main PISA Survey were composed of four clusters (two mathematics, one reading and one science). By using cluster rotation design, also known as balanced incomplete block design, 13 standard booklets for the Main PISA Survey were created. Main idea of this method is that each cluster (of 13) appear only in four test booklets (once in each of possible four positions in four-cluster booklet). Consequently, each student in different country was randomly assigned to the one of 13 booklets, hence total student testing time was considered to be about two hours (4 item clusters multiplied by 30 minutes of solving time).

The division of cognitive test items into clusters and usage of balanced incomplete block design ensured that all booklets can be linked together. This can be considered as one of the main features of the PISA survey, because it helps to take into account that booklets may have different difficulty levels. Therefore, PISA do not use raw student's score, but apply standardization procedure to neutralize the effect of test differences on final students scores. Other arguments also do not support the usage of raw scores. For example, student who gets zero correct answers will have a test score of 0 e but that doesn't mean he has no competencies; while student who gets all the correct answers cannot be considered as having all competencies [Wright and Stone, 1979].

In order to overcome those difficulties PISA applies Rasch Model for scaling. Detailed explanation of the item scaling theory and Rasch Model implementation can be found in PISA 2012 Technical Report [OECD, 2014] and PISA Data Analysis Manual SPSS® Second Edition [OECD, 2009].

3.3 Survey results

The outcome of survey, despite context student data are three major domains: mathematical literacy, reading literacy and scientific literacy. The definitions accent students' skills and functional knowledge about each domain, meaning that students are able to carry out more than just simple tasks that have one single correct answer, but to interpret or evaluate material. The definitions also focus on the students' ability to apply their knowledge in unfamiliar situations when mathematical concept of particular task is not obvious. Mathematical literacy (ML) has been defined as:

"An individual's capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals to recognize the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens" [OECD, 2013a, page 17].

Besides evaluating final mathematics score, known as ML, PISA also reports students' performance in content and processes areas. Content area in turn is divided into four traditional mathematics topics [OECD, 2016b]:

- Change and relationships. Tasks related to this topic often require from students the application of algebra and knowledge of mathematical models in order to describe and predict change.
- Space and shape. Tasks related to space and shape, compared with other topics most closely related to geometry. Usually tasks require from students' abilities to create and read maps, interpret three-dimensional scenes from different perspectives, transform and create representation of shapes.

- Quantity. Tasks related to quantity usually requires knowledge of numbers and number operations.
- Uncertainty and data: Tasks related to uncertainty and data has a strong connection with probability theory and statistics. Students usually need to demonstrate knowledge of variation in processes, uncertainty and error in measurement.

The processes area describes what students do and which methods apply to connect the context of particular task with their knowledge of mathematics and thereby solve a problem. PISA defines three processes – formulate, employ and interpret, which are drawn from a set of fundamental mathematical competencies required by students in order to use their mathematical knowledge. The brief explanation of those competencies is given below, while explicit definitions of such capabilities can be found in framework for [OECD, 2013a]:

- Communication. Ability of reading, decoding and interpreting statements with modeling follow-up.
- Mathematising. Transforming real world problems to a strict mathematical form.
- Representation. Ability to use a variety of available representations to interact with a problem.
- Reasoning and argument. The ability to use logical thinking, make inferences and provide justifications.
- Devising strategies for solving problems. The skill that helps students to create a plan or strategy how mathematics should be used in order to solve a problem.
- Using symbolic, formal and technical language and operations. The ability to use mathematics definitions, rules and formal systems described in symbolic expressions.

Considering possibility to evaluate not only the mathematical literacy of Lithuanian students, but a wide range of abilities defining mathematical skills, the work will analyze how different variables affect content and processes areas of mathematical knowledge.

3.4 Reported values

The “Reported values” subsection is hereby divided into three minor parts: plausible values, regressors and levels of analysis. The first part (3.4.1 on the next page) will give a brief explanation of the idea of plausible values (a dependent variables) that are used to report students score in PISA tests. Second part will present factors that are believed to have an impact on dependent variable, primarily e connected with mathematics plus some general ones. In third part potential levels of analysis will be observed and how they are defined by PISA.

3.4.1 Plausible values

PISA report student performance in mathematics (as well as other domains) through plausible values (PV). This methodology first time was implemented in the National Assessment of Educational Progress studies [Beaton, 1987]. The main motivation why this method is being used is that PISA belongs to the category of educational assessments with the purpose to assess the knowledge and skills of a population, rather that to measure knowledge and skills of individual. Consequently, the main goal is to reduce error when making inferences about the population, while minimization of measured error connected with each individual is not so important.

There are two reasons why PISA reports plausible values instead of students' actual final scores. First of all, to minimize the difference between reported value and true value, which can occur due to rounding process and measurement error. Secondly, to make inferences about unobserved proficiency levels of target population. As it was mentioned in the 3.2 on page 11, particular student final score will not necessarily represent his proficiency level in particular subject because of:

1. test complexity
2. mental and physical dispositions on the day of assessment
3. surroundings in which students are tested.

Methodology of plausible values consists of:

- mathematically computing distributions (denoted as posterior distributions) around the reported values; and
- assigning to each observation a set of random values drawn from the posterior distributions. In the particular PISA case – 5 plausible values are being assigned to each observation.

The meaning of plausible values and its methodology can be illustrated through example from sports. Imagine the situation that long jump (sport) tournament takes place, but it has a different format – the results of athletes are reported not as regular continuous variable, but in term of integers only, e.g. 3 meters, 4 meters, 5 meters and so on. In this case the range of possible results is a predefined number of values, while in reality observations can observed anywhere between minimum and maximum. Consequently, the jump length distribution can be drawn for those predefined numbers. So, if a value of 6 meters is taken it doesn't mean that reported jump length was exactly 6 meters, but only on average. Reader can remember already defined risks between reported and true value, which occur in this particular case. However, if the difference between the length of particular athlete's jump and closest integer is small, then probability of judge mistake is also small. For example, if the length of jump was equal to 6,05 meters, the reported value scarcely be 5 or 7 meters, but likely be 6 meters. Hereby, random values from the posterior distributions can be defined as plausible values [OECD, 2014]. From the particular example, a range of values from normal distributions around 6 meters (reported value) can be assigned to particular athlete's performance of 6,05 meters – e.g. 5,47 meters, 6,03 meters, 7,02 meters and so on.

The explanation of plausible values through sports example is connected with tests such as PISA, because it consists from dichotomous items – test tasks. The student raw score (e.g. number of correct answers) hereby is discontinuous variable, while his mental ability can be treated as continuous variable. It has been pointed out:

"The simplest way to describe plausible values is to say that plausible values are a representation of the range of abilities that a student might reasonably have. [...] Instead of directly estimating a student's ability ϑ, a probability distribution for a student's ϑ is estimated. That is, instead of obtaining a point estimate for ϑ, like a WLE, a range of possible values for a student's ϑ, with an associated probability for each of these values is estimated. Plausible values are random draws from this (estimated) distribution for a student's ϑ" [Adams and Wu, 2002].

Hereby, instead of one final ML score PISA data reports 5 plausible values of each test score (so for mathematics literacy) for each student. The same applies to sub-scales in mathematics, defined by processes and content areas. The main features of calculations using plausible values described in 4.3 on page 20. For more information about PV, see [Chapter 6 of PISA Data Analysis Manual: SPSS® Second Edition, OECD, 2009, pages 93-101] and [Chapter 9 of PISA 2012 Technical Report, OECD, 2014, pages 144-158].

3.4.2 Regressors

This part will describe the range of factors selected according to their connection with ML. The one section of those factors is directly connected with mathematics and its context, namely OTL variables, while other considered as general factors. Although the main focus of this study is to explore relationship between students' performance in mathematics and their knowledge of pure and applied mathematics, the importance of referred general factors cannot be ignored [Thien, 2016, Cogan et al., 2019]. Both exposure to pure mathematics and exposure to applied mathematics were measured by student-reported answers for a special questions, designed by PISA. Those indices were normalized to have average of 0 and standard deviation of 1 on OECD level. According to PISA 2012, OTL variables related to ML are (the abbreviations are defined in brackets):

- Exposure to pure mathematics (EXPUREM) – the experience with mathematical tasks that require knowledge of algebra, in particular – linear and quadratic equations, reported by students.
- Exposure to applied mathematics (EXAPPLM) – the experience of most commonly known "world" problems and straightforward tasks, such as calculating how many square meters of tiles you need to cover the floor or calculating a power consumption of electronic device per week, also reported by students.

Connected with mathematics and its context are:

- Familiarity with mathematical concepts (FAMCON) – students are being asked how familiar they are with different mathematical concepts. The question contained 16 different topics which included 3 non-existing pseudo-concepts with an eye to reveal over-claiming students. Consequently, the index was corrected and had same properties as exposure indices e average of 0 and standard deviation of 1.

- Interest in mathematics (INTMAT) – enjoyment of and interest in mathematics measured by student-reported answers.

- Learning time (MMINS) – the amount of time student spends in regular mathematical lessons (in minutes).

The general factors are:

- Grade (GRADE) – the actual grade of particular student in particular country whenever he participates in PISA assessment. Modal grade in Lithuania was 9 and values on individual level are reported as difference between student grade and modal grade. Important variable on within-country level because of the simple intuition – students in higher grade have had more years in education, consequently, they are more experienced and familiar with particular topic.

- Socio-economic status (ESCS) – is a broad concept that summarizes a variety of student, school and system factors. PISA estimates students' SES as an index based on "*such indicators as parents' education and occupation, the number and type of home possessions that are considered proxies for wealth, and the educational resources available at home*" [OECD, 2016b, page 74].

3.4.3 Levels of analysis

PISA results can be analyzed on between- and within- country levels. As it was mentioned before, the main focus of this work is to analyze Lithuania students' ML on different levels, which are motivated by cultural and political context. Therefore, the results of this work consist of:

- Country level. The students are considered as a unite population despite of differences on individual level.

- Area level. The PISA school questionnaire includes a question about the description of community in which particular school is located. School officials need to point out whether the school is located in[OECD, 2013a, page 213]:

  - A village or rural area (fewer than 3000 people)
  - A small town (from 3000 to about 15000 people)
  - A town (from 15 000 to about 100 000 people)
  - A city (from 100 000 to about 1 000 000 people)

- A large city (with over 1 000 000 people)

The particular level is being analyzed because of the context mentioned in introduction part – politicians in Lithuania underlines the inefficiency of schools located in villages and rural areas mainly because of the lack of funding and absence of high-quality teaching instructions (both are believed to be present in towns and cities). Hereby the work aims to test whether the exposure to pure and applied mathematics could be the key in order to narrow the gap between regions. PISA underlines that data derived from school questionnaires' should not be examined at school level but on student level with the use of student final and replicate weights.

- School type level. The stratification of school sample is used in order to improve the efficiency of sample design, namely by ensuring that all parts of population are included in the sample and each specific group of the population is adequately represented in the sample. The data set divides schools (and, as a consequence – students) into four groups:
  - Gymnasium
  - Secondary school
  - Basic school
  - Other

The motivation to analyze this level of stratification lies on the surface – it is believed that students from gymnasium perform better compared to other the students from other type of schools, especially basic and others, which often includes the schools of national minorities, in Lithuania. The reasons of such beliefs are closely related to area level stratification – the better schools (consequently with higher "status") are located in the "crowded" areas with sufficient funding, while smaller schools may experience student and funding scarcity. The funding argument is also connected with quality of teaching e dependent variable of teacher's well-being. Hereby, more motivated and qualified teachers tend to choose better schools, while basic and others schools may experience scarcity for such teachers.

- SES level. At this level students are categorized according to their socioeconomic profile. The importance of SES was stated by numerous authors[Baker et al., 2002, Breen and Jonsson, 2005, Chiu and Khoo, 2005] and PISA all-time reports. In this work author uses three levels of SES described by PISA[OECD, 2016b, page 33] :
  - Socioeconomically advantaged students. The students whose economic, social and cultural status (ESCS) index value is at or above the 75th percentile of their own country.
  - Socioeconomically disadvantaged students. Consequently, the students whose ESCS index value is below the 25th percentile in their own country.
  - Socioeconomically "neutral" students. The students whose ESCS index value is between 25th and 75th percentiles of their own economy.

- Proficiency level. Proficiency scales have been developed by PISA in order to make results more accessible to governments. The proficiency level can be understood as a composition of skills required by student to successfully solve problems with particular difficulty level. They are not included in final PISA data sets, but can be derived from the plausible values and cut-points reported by particular year report. The total of 7 levels can be derived from the date, starting from below one (lowest possible) and ending with 6th (highest possible). For more information about proficiency scale descriptions, see [OECD, 2013a, PISA 2012 Framework, p.41]. Instead of analyzing each group of students according to proficiency levels, the author will focus on low, normal and top performers:
  - Low performers – students at or below proficiency Level 1, who scored less than 420.1 points.
  - Top performers – students at Level 5 or 6 of proficiency, who scored more than 607 points.
  - Normal (average) performers – students between 2 and 4 proficiency level, who scored from 420.1 to less than 607.

4 Methodology

This part will introduce specification of models that will be used in the analysis, the implementation of weights in multilevel models and general ideas of statistic analysis with plausible values.

4.1 Specification of various models

The fact that PISA has two-stage stratified sample design implies the potential to use multilevel models (so-called models with hierarchical structure, where it appears as it described in 3.1 on page 9 part). As a consequence, such model with PISA survey data will have two levels:

- Level-1 e students (units i from second stage of sample design);
- Level-2 e schools (units j from first stage of sample design).

The main idea why levels appear and are important in analysis is the reason that the assumption of independence does not always hold in two-stage design [Raudenbush and Bryk, 2002]. The multilevel model takes into account that students are nested within schools and classes. As schools are first stage of stratified sampling (and the level-2 of multilevel model) the students within same school may have more in common compared to students from another schools[OECD, 2009]. Therefore, the assumption of independence should be adapted by using variance component models which decompose the variance into group and individual parts: group components are perfectly correlated within groups, but independent between; individual components are all independent [Raudenbush and Bryk, 2002].

With respect to two-stage PISA survey data design the mathematical specification of model has the form:

$$y = X\beta + Zu + \varepsilon \tag{2}$$

where y is the vector of outcomes, X is a matrix of covariates associated with regressors that are assumed to be fixed, β is the vector of fixed-effect regression coefficients, Z is a matrix of covariates associated with regressors that are assumed to be random, and u is the vector of random effects. The models usually fitted by PISA [OECD, 2009] and other authors [Chiu and Khoo, 2005, Perry and McConney, 2010, Cogan et al., 2019] are type of multilevel models called Hierarchical Linear Models (HLM) with two levels and following notation and form[Raudenbush and Bryk, 2002]:

$$y_{ij} = \beta_{0j} + X\beta_{1j} + \varepsilon_{ij} \tag{3}$$

$$\beta_{0j} = \gamma_{00} + \delta_{0j} \tag{4}$$

$$\beta_{1j} = \gamma_{01} + \delta_{1j} \tag{5}$$

Where y_{ij} is the first level, β_{0j} and β_{1j} are second levels, random intercept and random slope models respectively; δ_{0j} and δ_{1j} are the error terms for the intercept and slope and have variances of τ_{00} and τ_{11}. The efficiency of multilevel model can be calculated using Intraclass Correlation parameter:

$$IC = \frac{\tau_{00}}{\tau_{00} + \tau_{11}} \tag{6}$$

If $\delta_{0j} = \delta_{1j} = 0$, so Level-2 variances $\tau_{00}$and $\tau_{11}$are equal to 0 and HLM would be mathematically equal to a simple linear regression. The differences between those two model tend to be bigger if Level-2 variance gets bigger, so multilevel model produce more precise estimates of standard errors, while linear regression underestimates them. However, the motivation to use multilevel model over linear regression directly connected with the social segregation in particular country, more precisely e differences between schools. E.g., if some schools in country are attended primarily by high-SES students, while others by low-SES e such segregation will create social segregation, so if y_{ij} is correlated with SES, multilevel model will give a better fit.

Considering the availability of sub-scores of ML described in 3.3 on page 12, the multivariate multiple regression (MMR) will be used. This technique expands opportunities of multiple linear regression by including the ability to handle multiple outcome variables:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \tag{7}$$

where **Y** is a matrix of k observations on n dependent variables, **X** is a matrix for independent variables, **B** is a matrix of regression coefficients and **E** is a matrix of errors. Therefore, besides primary analysis on students general mathematics score, the n will consist of additional 7 sub-scores in mathematics (content and processes areas).

4.2 Weights

For the most analyses it is recommended and enough to use student final weight W_{ij}, while multilevel analysis cannot be performed only with W_{ij}, especially if the assumption of equal probability sampling at the first stage does not hold (as it is in PISA). Therefore the multilevel analysis suggests the usage of conditional weights retrieved from W_{ij} or the usage of weight rescaling procedure in order to make w_{2ij} independent from w_{1i} [Rabe-Hesketh and Skrondal, 2006]. PISA reports w_{1i}, so the conditional weight w_{2ij} can be obtained as:

$$w_{2ij} = W_{ij}/w_{1i} \tag{8}$$

When w_{2ij} is available, rescaling becomes available, but still important decision to make, because usage of weight rescaling methods can help to avoid biased estimates. Notably weight scaling methods should be used when number of students within a school is small and students base weights (Level-1) are very different from 1 [Rabe-Hesketh and Skrondal, 2006]. The scaling method used in this work is the one described in[Pfeffermann et al., 1998]:

$$W_{ij}^{*} = W_{ij}\left(\frac{n_i}{\sum_j W_{ij}}\right) \tag{9}$$

where j indexes the individuals, i e the groups and n_i represents number of students in group i.

The motivation to use this scaling method comes from simulations which suggest that it works better for informative weights [Pfeffermann et al., 1998]. This method is also defined as default for PISA survey data in the *EdSurvey* package of *R* statistical software. With weighting being a complex issue in recent years[2], in this work author will consider analysis with both weight transformation methods as a form of sensitivity analysis.

4.3 Analysis with plausible values

As it was stated in 3.4.1 on page 14, PISA reports 5 plausible values for each particular student score. The "correct" way of calculating population statistics is to estimate particular statistic θ using each plausible value separately and then average them. If N is number of plausible values, then mathematically:

$$\lambda = \frac{1}{N}\sum_{i=1}^{N}\lambda_i \tag{10}$$

Then the final error variance equals to the sum of sampling variance and imputation variance[OECD, 2014, page 148]:

[2]PISA has been using weight "normalization" procedure so the sum of the weights is equal to the number of students in the dataset. However, the use of separate weights at different levels was under technical discussion, as it stated in [OECD, 2009, page 222]

$$V = U^* + (1 + N^{-1})B_N \tag{11}$$

where sampling variance U^* is the average of N sampling variances

$$U^* = \frac{1}{N}\sum_{i=1}^{N} U_N \tag{12}$$

where U_N is a sampling variance calculated using a replication methodology known as Balanced Repeated Replication (BRR), Fay's method [OECD, 2014]. In countries, where not all schools were selected, but a sample of schools, it was decided (by PISA) to generate 80 replicate weights[OECD, 2009]. Sampling variance formula then becomes:

$$U_N = 1\frac{1}{G(1-k)^2}\sum_{i=1}^{G}(\hat{\lambda}_i - \hat{\lambda})^2 \tag{13}$$

where G = 80 and k = 0.5 (Fay method's factor).

B_N is imputation variance equals to:

$$B_N = \frac{1}{N-1}\sum_{i=1}^{N}(\lambda_i - \lambda)^2 \tag{14}$$

Therefore, final standard error is equal to:

$$SE = \sqrt{V} \tag{15}$$

5 Practical part

5.1 Exploratory analysis

To start with, general statistics are derived from the data. According to PISA 2012 data, average mathematics score of Lithuania students was 478.82. Sample size of 4618 students represented population of total 33042 students. In the Table 1 on page 23 we can see sample size, weighted sample size, mean score, population percentages and standard error of mean and percentage values in Lithuania and its sub-levels (region, type of school, SES group, grade). Additionally, "D.mean" column show the difference between country mean and particular sub-level mean, where statistically significant ($p < 0.05$) differences are **bold**.

Lithuanian students from schools located in cities score the most on average compared with other regions. The 37.49 per cent of total population have the average result near 500 points and the difference with total population is statistically significant. However, there are no statistically significant difference between students from schools located in the cities and towns, only with small towns and villages. The one and only region which has statistically significant mean differences compared with others e village. Such results confirm that there are differences between cities and rest of the regions, just as it was mentioned in 1 section.

More detailed look on school type level confirms that students from gymnasium score the most. Raw mean differences with other school types vary between 32.44 e 72,02 points. Such results follow the consistent intuition that gymnasiums usually have better quality of teaching with combination of other factors (mainly located in cities, have students with higher SES). However, there are no statistically significant differences between "other" and the rest types of schools. It can be explained by a vast variation of in mean scores between schools of "other" type, likely because this type includes special or (and) ethnic minorities schools.

Socioeconomically advantaged students on average score 521.57 points. The differences between this type of students and others (with average SES and SES disadvantaged) are all statistically significant. As it was mentioned 2 section, SES is one of the main factors which affect students knowledge and test scores.

The last level of analysis is grade. The target population of PISA (15-years-old students) in Lithuania mainly (81.21 per cent) were in the 9th grade (out of twelve) with the average score of 480.01 points. The results from Table 1 confirms the PISA test scores varies a lot between depending on student grade. E.g., students from 11th grade scored 597.86 points, while students from 7th grade only 347.82 points. Such differences suggest the importance of grade factor and the fact that analysis must control this effect.

The values in the table are calculated using formulas described in 4.3 on page 20 with unconditional final student weight. Graphical representation of mean differences are presented in Figure (5) e Figure (8) on pages 36e38.

Level	N	Weighted N	Mean	SE(Mean)	Percent	SE(Percent)	D.mean
Lithuania	4618	33041.96	478.82	2.64	100.00	0.00	0
City	1736	12387.33	498.74	3.73	37.49	1.05	**-19.92**
Town	950	6837.43	488.83	8.14	20.69	2.41	-10.01
Small Town	1042	7230.43	469.09	6.00	21.88	2.43	9.73
Village	890	6586.77	441.66	4.37	19.93	1.25	**37.16**
Gymnasium	2715	19182.73	499.69	3.21	58.06	0.77	**-20.87**
Secondary	1052	7523.57	467.25	4.93	22.77	0.64	**11.57**
Basic	746	5687.85	429.58	5.95	17.21	0.62	**49.24**
Other	105	647.81	427.67	39.88	1.96	0.69	51.15
SES advantaged	1150	8208.74	521.57	3.32	25.03	0.93	**-42.75**
SES avg	2283	16350.17	478.05	2.73	49.86	0.92	0.77
SES disadvantaged	1148	8230.37	438.90	3.50	25.10	0.88	**39.92**
Modal grade + 2	2	13.38	597.86	18.39	0.04	0.03	**-119.04**
Modal grade + 1	525	4060.72	506.91	5.96	12.36	0.72	**-28.09**
Modal grade	3761	26680.45	480.01	2.63	81.21	0.69	-1.19
Modal grade e 1	278	2049.27	418.76	6.35	6.24	0.55	**60.06**
Modal grade e 2	7	50.17	347.82	21.99	0.15	0.06	**131.00**

Table 1: Descriptive statistics of ML in Lithuania.

Another possible way of grouping students can be derived using proficiency scales in ML. The detailed results with cut-points, sample and population number of students and percentages of students on different proficiency levels are presented in Table 2. According to PISA 2012, 26.01 per cent of students in Lithuania do not achieve second level of proficiency in mathematics, and only 8.06 per cent of students are ranked as top performers. Both shares of students are statistically different compared with OECD average with share of low performers above and share of top performers below OECD average.

Level	Cut-points	N	Weighted N	Percent	SE(Percent)
Below Proficiency Level 1	(<) 357.77	395.00	2889.27	8.74	0.68
At Proficiency Level 1	420.07	799.20	5707.69	17.27	0.89
At Proficiency Level 2	482.38	1201.00	8569.80	25.94	0.80
At Proficiency Level 3	544.68	1136.00	8128.27	24.60	1.01
At Proficiency Level 4	606.99	711.80	5083.39	15.38	0.70
At Proficiency Level 5	669.3	307.60	2188.86	6.62	0.49
At Proficiency Level 6	(>) 669.3	67.40	474.69	1.44	0.24

Table 2: Share of students by proficiency level

5.2 Modeling step

While final student weights can be used in almost all analysis, the chosen specification of model (multilevel model) implies the use of different weights at different levels. However, PISA do not report higher stages weights, e.g. regions, so in this work author consider multilevel model with two levels according to two-stage sampling design e schools and students. As it was mentioned in 4.1 on page 18, the multilevel model can be considered as a mix of fixed and random effects. Bearing in mind that Level-2 weights are accessible, the choice between fix and random effects in multilevel model is present.

As the first step of explanatory analysis of possible multilevel model specifications, differences in ML mean scores between schools in Lithuania were checked. The aim of this step is to analyze what kind of intercept effect must be selected. Multilevel model with fixed intercept will have the same intercept as linear regression model, if δ_{oj}in 4 on page 19 will be equal to zero. The opposite site, model with random intercept will result in random intercept for each Level-2 unit (school). In the Figure 1 we can see that schools in Lithuania vary a lot in ML mean scores.

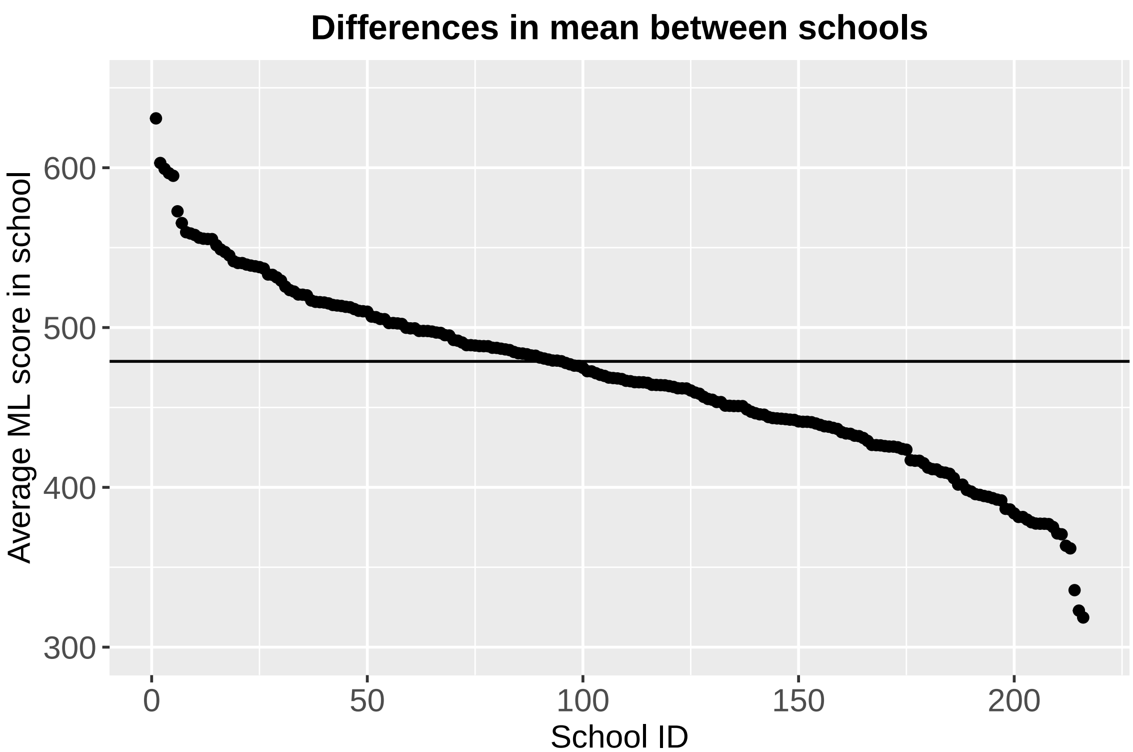

Figure 1: Mean differences of ML scores in Lithuania by school ID

This is an important argument in order to use multilevel model with random intercepts at Level-2 (school level). It is also suggested by [OECD, 2009] to analyze decomposition of variance of the dependent variable which is divided into the within-school and between-school variance. Hereby, the multilevel model formula 3 on page 19 equals to:

$$y_{ij} = \beta_{0j} + \varepsilon_{ij} \tag{16}$$

$$\beta_{0j} = \gamma_{00} + \delta_{0j} \tag{17}$$

The amount of variation explained by school differences are presented in the Table (3), where

6 possible outcomes are presented, depending on used number of observations due to missing data. outcome 1-2 e full population, outcome 3-4 e after controlling for SES and grade, outcome 5-6 e after controlling for SES, grade and other variables related to ML. There are differences in results when two types of weights were used: weights adjusted by scaling method (scaled W) seem to give much lower between-school variance estimates with bigger intercepts, while conditional weights (conditional W) imply bigger between-school variance, but smaller within-school variance, so the intraclass correlation becomes bigger. The on should bear in mind that model with conditional weights may result in biased estimates, while scaling method could over-correct bias. Hereby by, both situations will be considered to analyze whether final model estimates are sensitive to weights.

Models with weights	**N**	**Between-school variance**	**Within-school variance**	**IC**	**Intercept**
Outcome 1, scaled W	4618	1429.331	6139.368	0.189	462.327
Outcome 2, conditional W	4618	2372.985	5341.795	0.308	451.207
Outcome 3, scaled W	2987	1203.030	6348.226	0.159	467.847
Outcome 4, conditional W	2987	2482.872	5204.127	0.323	454.253
Outcome 5, scaled W	1434	584.9595	7124.110	0.076	479.529
Outcome 6, conditional W	1434	2556.737	5111.529	0.333	460.368

Table 3: “Empty” model with random intercepts between- and within-school variance estimates and intraclass correlations

As the second step, the choice between fixed and random slopes is considered. The combination of random intercepts and fixed slopes will result in a model that is very similar to ordinary least squares (OLS) method, however starting point of each slope will vary depending on school intercept. A multilevel model with random intercepts and slopes would mean that each Level-2 factor (school) will have both unique intercept and slope. The intuition of multilevel model is also different: while OLS procedure show the exact value of intercept and slope, multilevel model with both random effects shows the average value of the intercept and slope between chosen schools. Such model specification can reasonably reduce variation, however the overall reduction depends on the differences between schools which appear when regressors from 3.4.2 on page 15 are under analysis. Figure 2 on the following page suggest that there are not so much differences in exposure to pure mathematics slopes between schools. Initial analysis of three main factors (student SES, exposure to pure and applied mathematics indices) show that slopes hardly vary between different regions and different schools. Graphical representation of random slopes for sub-populations presented in Figure (9) e Figure (11) on pages 38e39.

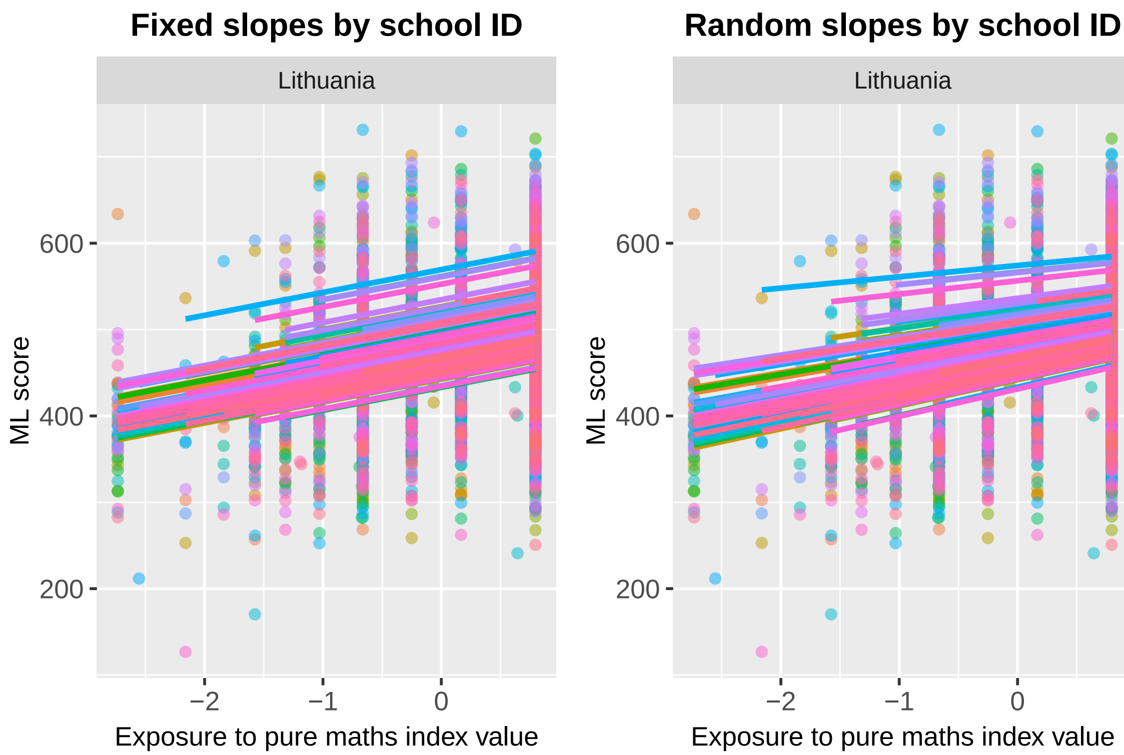

Figure 2: Random intercepts with fixed and random slopes for exposure to pure mathematics index value by school ID

Practical recommendation when fitting multilevel models is to compare estimates the with estimates from OLS procedure. Another important argument in order to perform OLS procedure e the fact that it is simpler to explain. Hereby, next part will examine results from 3 different models: random intercept model with fixed slopes, random intercept model with random slopes and linear regression. It will result in 5 different model outcomes, because unconditional weights and scaling procedure will be used to examine whether the multilevel model estimates are sensitive to weights (both multilevel models will be fitted using two different weight techniques). The fifth outcome will be OLS estimates (with final student weight only). So, the equation of model with four primary Level-1 independent variables will have a form:

$$y_{ij} = \beta_{0j} + \beta_{1j}(GRADE)_{ij} + \beta_{2j}(ESCS)_{ij} + \beta_{3j}(EXPUREM)_{ij} + \beta_{4j}(EXAPPLM)_{ij} + \varepsilon_{ij} \tag{18}$$

When multilevel model with random intercept and fixed slopes is considered, the random intercept will have form defined in Equation 4 on page 19, while multilevel model with both random intercept and random slopes, random effects β_{ij} will be equal to:

$$\beta_{1j} = \gamma_{10} + \delta_{1j} \tag{19}$$

$$\beta_{2j} = \gamma_{20} + \delta_{2j} \tag{20}$$

$$\beta_{3j} = \gamma_{30} + \delta_{3j} \tag{21}$$

$$\beta_{4j} = \gamma_{40} + \delta_{4j} \tag{22}$$

For model with more Level-1 independent variables (MMINS, FAMCON, INTMAT from Subsection (3.4.2)) same logic will be applied.

In the next section results of described models are presented using abbreviations:

- Multilevel models with random intercepts and fixed slopes are defined as **Model 1** and **Model 2**, where Model 1 was fitted using weight scaling procedure, whereas Model 2 e using conditional weighs.
- Multilevel models with both random intercepts and slopes are defined as **Model 3** and **Model 4**, where Model 3 was fitted using weight scaling procedure, whereas Model 4 e using conditional weighs.
- The linear regression model is defined as **Model 5**.

6 Results

The first subsection will discuss results of analysis for Lithuania: present different models and overview relationship of variables defined in hypothesis with the mathematics literacy. In the second part results will be compared with PISA reports and results from [Cogan et al., 2019]. In addition, some of the models will be fitted for Latvia and Estonia in order to check whether hypothesis holds for neighboring countries and what are the differences in estimates.

6.1 A closer look at the Lithuania: results from different models

The estimates of intercept, grade, SES index, exposure to pure and applied mathematics indices of 4 different model specification with 5 outcomes portrayed in Table 4 on the following page. All estimates in 5 models are statistically significant ($p<0.01$), except index of exposure to applied mathematics e only in one model it is statistically significant with at $p = 0.05$ and in other two cases with at $p = 0.1$. Three predictors out of 4 have positive impact on ML score e grade, SES and exposure to pure mathematics index, while exposure to applied mathematics index has a negative sign. However, it is clear that estimates vary between models, especially between multilevel models with same specification but different weights technique applied. A closer look on IC also confirms differences in identical models with different weights, e.g. Model 1 and Model 2 (random intercept multilevel model with fixed slopes) have IC equal to 0.074 and 0.259, meaning that in first case specified predictors explain only 7.4 per cent of variation between-schools, while the rest 92.6 percent is due to within-schools variation, while in the second e 25.9 per cent of variation between-schools and 74.1 percent within.

Even greater difference in IC appear in multilevel model with both random intercept and slope. In this case, Model 4 with unconditional weights explains 50.1 per cent of variation on between-school level, while Model 3 with weights scaling procedure indicates that there are only 8.4 per cent of variation explained by differences in schools.

Effect	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Intercept	474.54*** (2.86)	460.67*** (3.77)	474.08*** (2.86)	463.34*** (3.86)	479.42*** (2.17)
Grade	31.52*** (3.46)	31.99*** (3.42)	26.23*** (1.79)	19.01*** (1.81)	29.81*** (4.35)
Socioeconomic status index [escs]	27.03*** (1.80)	18.14*** (1.61)	31.25*** (3.44)	30.03*** (3.67)	31.47*** (1.88)
Exposure to pure mathematics index [expurem]	23.96*** (1.93)	18.33*** (1.84)	24.41*** (1.93)	19.10*** (2.10)	26.70*** (2.13)
Exposure to applied mathematics index [exapplm]	-3.71 . (1.96)	-2.02 (1.87)	-3.96* (1.95)	-3.23 (2.24)	-3.74 . (2.14)
Model statistic	IC	IC	IC	IC	R^2
Value	0.074	0.259	0.084	0.501	0.221

Note: Standard errors are in parentheses. Predictors abbreviations are in square brackets.
Significance codes: .p < .1. *p < .05. **p < .01. ***p < .001.

Table 4: Results from 5 different models with primary predictors

The efficiency of multilevel model can also be checked by comparing the difference in explained variance between fitted model and empty model (with only random intercepts). Hereby, comparison between Table 3 on page 25 and Table 5 on the following page is considered. As we can see, Model 1 between-school residual variation decreased from 1203.030 to 446.207, while within-school residual variance decreased from 6348.226 to 5572.540. Therefore, the amount of variance explained by *grade*, *escs*, *expurem* and *exapplm* at school level is equal to $1 - \frac{446.207}{1203.030} = 0.629$, while at the student level: $1 - \frac{5572.540}{6348.226} = 0.122$. Such results suggest that there are some segregation in schools and model catches them. The common thinking suggest that those differences appear mainly because of the differences in SES status (so one part of school is being attended by SES-advantaged students, while other e by disadvantaged. The same inferences can be applied to Model 3 with random slopes. Another interesting case appear when conditional weights are being used. So while Model 2 with random intercepts and fixed slopes follows the familiar path and decreases between- and within-school variance, the Model 4 with both random effects substantially increases school variance : from 2482.87 to 3876.47. It can be explained by the fact the lower-performing students may have deeper slopes while high-achievers e flatter.

Models with weights	N	Between-school variance	Within-school variance	IC
Model 1, scaled weights	2987	446.207	5572.540	0.074
Model 2, conditional weights	2987	1620.288	4638.146	0.259
Model 3, scaled weights	2987	506.120	5551.081	0.084
Model 4, conditional weights	2987	3876.466	3862.261	0.501

Table 5: Between- and within-school variance estimates and intraclass correlations in multilevel model with random intercepts and fixed slopes

Usage of two techniques to apply weighted analysis, as a part of sensitivity analysis suggest that estimates are stable in meaning of statistical significance and sign of relationship. Model 1 suggest lower standard error at school level (intercept) compared to Model 2, but larger standard error at student level, while Model 2 suggest smaller standard errors for all predictors. Model 3 suggest lower standard errors for all estimates compared to Model 4. However, according to [Rabe-Hesketh and Skrondal, 2006], while unconditional (raw) Level-1 weights may produce positive bias, scaling methods may over-correct it in smaller cluster sizes.

Another part of sensitivity analysis is to compare multilevel model results with estimates from OLS procedure. If scaling procedure is considered as primary, the estimates linear regression are very similar to estimates from Model 1 and Model 3. However, if no scaling method applied, Intraclass Correlation from Model 2 and Model 4 suggest that multilevel models should be used. However, for the hypothesis testing procedure there is no need to choose one model, because there are no differences in signs of relationships and significance level between multilevel models themselves and compared to OLS procedure.

Exposure to pure mathematics has a positive relationship and is statistically significant ($p < 0.001$) in all sub-scores of mathematical knowledge, whether exposure to applied mathematics has a negative relationship and is statistically significant ($p < 0.01$) in "uncertainty and data" and "interpret" sub-scores. The results of multivariate regression for ML sub-scores are presented in Table (6). It is clear that results of multivariate regression are very similar because sub-scores correlate with each other as well as with final ML score (residual correlation matrix can be examined in Table 9 on page 40).

Initial model with primary predictors supports the hypothesis that exposure to pure mathematics has bigger (and positive) impact on ML scores, compared with exposure to applied mathematics index. This statement holds for all 5 defined models even after controlling for student grade and SES index, also for all sub-scores (content and processes areas) of mathematical knowledge. While relationship of exposure to pure mathematics to ML seem to have a strong link in all cases, the relationship of exposure to applied mathematics and ML is questionable. The main concern here is the negative sign e meaning that student with higher exposure to applied mathematics index value will have lower ML results. A deeper look at proficiency levels suggest that exposure to applied mathematics has a positive relationship with ML, but after a certain level of exposure to applied mathematics is reached ML starts to decrease. The quadratic relationship of applied mathematics to ML was analyzed by [Cogan et al.,

Effect	Change and Relation-ship	Quantity	Space and shape	Uncertain-Ty and data	Employ	Formulate	Interpret
Intercept	479.96*** (2.66)	482.93*** (2.45)	473.72*** (2.64)	473.79*** (2.33)	482.76*** (2.30)	477.76*** (2.61)	473.10*** (2.64)
Grade	32.64*** (4.71)	28.34*** (4.82)	29.66*** (4.82)	31.26*** (3.92)	27.37*** (4.13)	31.71*** (5.05)	25.84*** (4.52)
Socio-economic status index	32.02*** (2.06)	31.40*** (2.06)	32.77*** (2.52)	30.66*** (1.96)	29.86*** (1.98)	34.78*** (2.37)	31.98*** (1.99)
Index of exposure to pure mathematics	26.79*** (2.24)	27.41*** (2.36)	27.26*** (2.42)	26.86*** (2.06)	26.40*** (2.25)	29.03*** (2.71)	24.13*** (2.26)
Index of exposure to applied mathematics	-4.15 . (2.16)	-2.23 (2.37)	-3.70 (2.26)	-5.63** (2.22)	-2.14 (2.16)	-3.73 (2.24)	-6.39** (2.26)
R^2	0.22	0.20	0.19	0.21	0.22	0.20	0.19

Note: Standard errors are in parentheses.
Significance codes: .p < .1. *p < .05. **p < .01. ***p < .001.

Table 6: Multivariate regression results on different ML sub-scores

2019], however, no analysis for Lithuania was performed. The graphical representation of both indices relationships to ML show that exposure to applied mathematics has almost similar form to quadratic. It can be seen in 3 on the following page that despite of low starting point (below 400) when index of exposure to applied mathematics reaches -1, the illustrated line breaks the average ML score in Lithuania. However, after this point there are no significant increase in ML despite growth of exposure to applied mathematics, whereas ML is increasing as exposure to pure mathematics increases.

The [OECD, 2016b] suggest that there are significant relationship between other factors connected with mathematics (familiarity with concepts, interest in mathematics and learning time per week) and ML. Hereby, the Equation 3 on page 19 will have a form:

$$y_{ij} = \beta_{0j} + \beta_{1j}(GRADE)_{ij} + \beta_{2j}(ESCS)_{ij} + \beta_{3j}(EXPUREM)_{ij} + \beta_{4j}(EXAPPLM)_{ij} + \beta_{5j}(MMINS)_{ij} + \beta_{6j}(FAMCON)_{ij} + \beta_{7j}(INTMAT) + \varepsilon_{ij}$$

The results of such model specification are shown in Table 7 on page 32. However there are only two model specifications: multilevel model with random intercepts and fixed slopes and linear regression. The specification of multilevel model with random slopes is inaccessible because there are not enough observations for generated amount of random factors. According to the Table 7 on page 32, statistically significant relationship with ML have 4 predictors: student SES, exposure to pure mathematics index, familiarity with mathematical concepts and interest in mathematics, whether grade, exposure to applied mathematics and learning time per week aren't significant. The IC of Model 1.1 is very small (0.0009), while IC of Model 2.2 equals to 0.241. The differences in IC between model with same specification appear due to differences in weights. In addition, Model 1.1 have similar estimates to Model 5.1. Even after controlling for more predictors exposure to pure

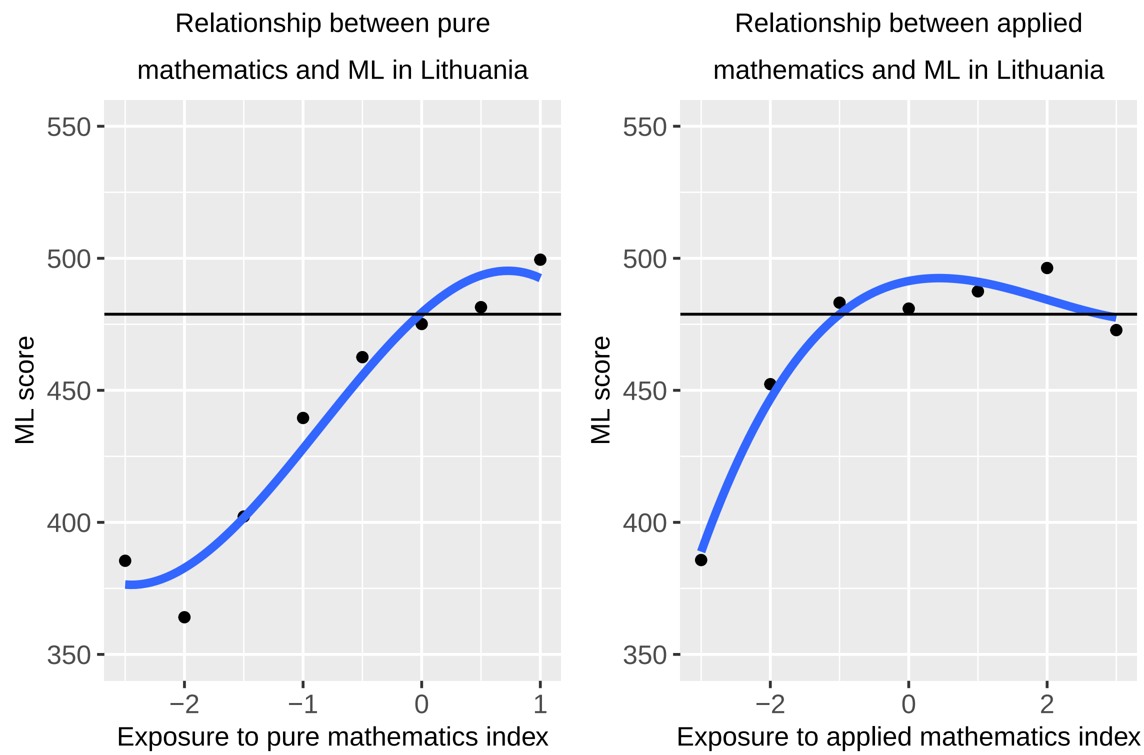

Figure 3: Relationship between pure and applied mathematics indices and ML in Lithuania

mathematics has a strong link with ML, being statistically significant ($p < 0.001$) in all three model outcomes, whereas exposure to applied mathematics being insignificant ($p < 0.1$) in all three cases.

Effect	**Model 1.1**	**Model 2.1**	**Model 5.1**
Intercept	459.33*** (16.10)	481.95*** (9.38)	459.48*** (37.51)
Modal grade -1	-23.45 (15.38)	-35.55 (7.71)	-23.94 (35.44)
Modal grade -1	24.11 (13.49)	-1.60 (5.17)	23.13 (35.01)
Modal grade +1	27.71 (15.32)	9.81 (8.57)	23.49 (35.66)
Socioeconomic status index [escs]	26.15*** (2.27)	17.82*** (2.12)	25.83*** (2.30)
Exposure to pure mathematics [expurem]	20.34*** (2.68)	16.49*** (2.47)	21.02*** (2.89)
Exposure to applied mathematics [exapplm]	-4.76 . (2.65)	-3.91 . (2.66)	-4.48 . (2.57)
Learning time per week [mmins]	0.03 (0.05)	-0.03 (0.04)	0.04 (0.05)
Familiarity with mathematical concepts [famcon]	31.66*** (3.23)	28.33*** (3.09)	32.44*** (3.48)
Interest in mathematics [intmat]	7.85*** (1.89)	9.16*** (1.76)	7.50*** (2.03)
Statistic	**IC**	**IC**	R^2
Value	0.0009	0.241	0.3315

Note: Standard errors are in parentheses. Predictors abbreviations are in square brackets.
Significance codes: .p < .1. *p < .05. **p < .01. ***p < .001.

Table 7: Results from 3 different models with all related to mathematics predictors

6.2 Comparison with other countries

In general PISA data show that students' ML scores positively associated with their exposure to pure and applied mathematics. The [Figures 3.8a in OECD, 2016b, p. 139] show whenever students from different countries are more frequently exposed to pure mathematics problems, their performance in mathematics grows. According to this analysis a one unit increase in index of exposure to pure mathematics will result in more than 30 points positive score-difference for student in Lithuania. Those results do not contradict the estimates from 6.1 on page 27, where index of exposure to pure mathematics is associated with 18-27 points increase in student ML score. Comparing with other countries, a one unit increase in index of exposure to pure mathematics for average student will result in 3-47 points increase in performance in mathematics. However, more frequent exposure to applied mathematics do not always associate with increase of ML. On average one unit increase in index of applied mathematics will increase ML score by about 9 points [OECD, 2016b, p. 138]. However, in some countries this relationship is negative. According to PISA, Lithuania has positive relationship between

Effect	Model 1 LT	Model 1 LV	Model 1 EST	Model 5 LT	Model 5 LV	Model 5 EST
(Intercept)	474.54*** (2.86)	505.52*** (2.42)	525.64*** (2.42)	479.42*** (2.17)	507.34*** (2.23)	525.51*** (2.35)
Grade	31.52*** (3.46)	48.48*** (3.43)	35.08*** (2.80)	29.81*** (2.86)	48.47*** (2.86)	32.92*** (2.99)
Socioeconomic status index [escs]	27.03*** (1.80)	24.14*** (1.63)	20.58*** (1.91)	31.47*** (1.88)	26.58*** (2.86)	25.40*** (1.90)
Exposure to applied mathematics [exapplm]	-3.71 . (1.96)	-1.98 (1.72)	-4.59** (1.78)	-3.74 (2.13)	-2.90 (2.19)	-1.97 (2.01)
Exposure to pure mathematics [expurem]	23.96*** (1.93)	17.86*** (1.77)	12.55*** (1.42)	26.70*** (2.14)	19.05*** (1.95)	13.19*** (1.73)
Statistics	**IC**	**IC**	**IC**	R^2	R^2	R^2
Value	0.074	0.076	0.097	0.221	0.284	0.141

Note: Standard errors are in parentheses. Predictors abbreviations are in square brackets.
Significance codes: .p < .1. *p < .05. **p < .01. ***p < .001.

Table 8: Results from two models using primary predictors

index of applied mathematics and ML, but the analysis in precious part suggest that this relationship has a quadratic form, therefore to a certain level relationship is positive, while after crossing that level it becomes negative. The work of [Cogan et al., 2019] show that "School maths" (an OTL variable which is closely related to exposure to pure mathematics) have strong link with performance in mathematics, while "Applied maths" primarily had a quadratic form. Therefore, results from different models analyzed in this work at Lithuania level do not disagree with conclusion made by those authors.

In addition the results from 6.1 on page 27 for Lithuania were compared with Latvia and Estonia using same models. At the first step the differences between-schools were checked using empty model with only random intercept at Level-2. The IC values of Latvia and Estonia equal to 0.166 and 0.133, while Lithuania had IC equal 0.189 using weights scaling technique and full available data set. The analysis of Lithuania using different models showed that predictors are resilient to various weights techniques as well as models. Therefore, comparison of Lithuania with Latvia and Estonia will be made using multilevel model with random intercepts and fixed slopes (Model 1); linear regression (Model 5). Result are shown in 8

Latvia and Estonia statistically significant estimates of the models are the same, as in Lithuania: the one variable, which is not statistically significant in all models e index of exposure to applied mathematics. It is important to note, that in PISA 2012 Latvia and Estonia had bigger mean scores in mathematics compared to Lithuania (LV e 491 points, EST e 521 points). That's why Lithuania intercept the smallest compared to neighboring countries. Grade variable in Latvia and Estonia has greater impact on results compared to Lithuania. Estimate of SES index suggest that Lithuanian students' performance in mathematics are affected by their SES more that in neighboring countries.

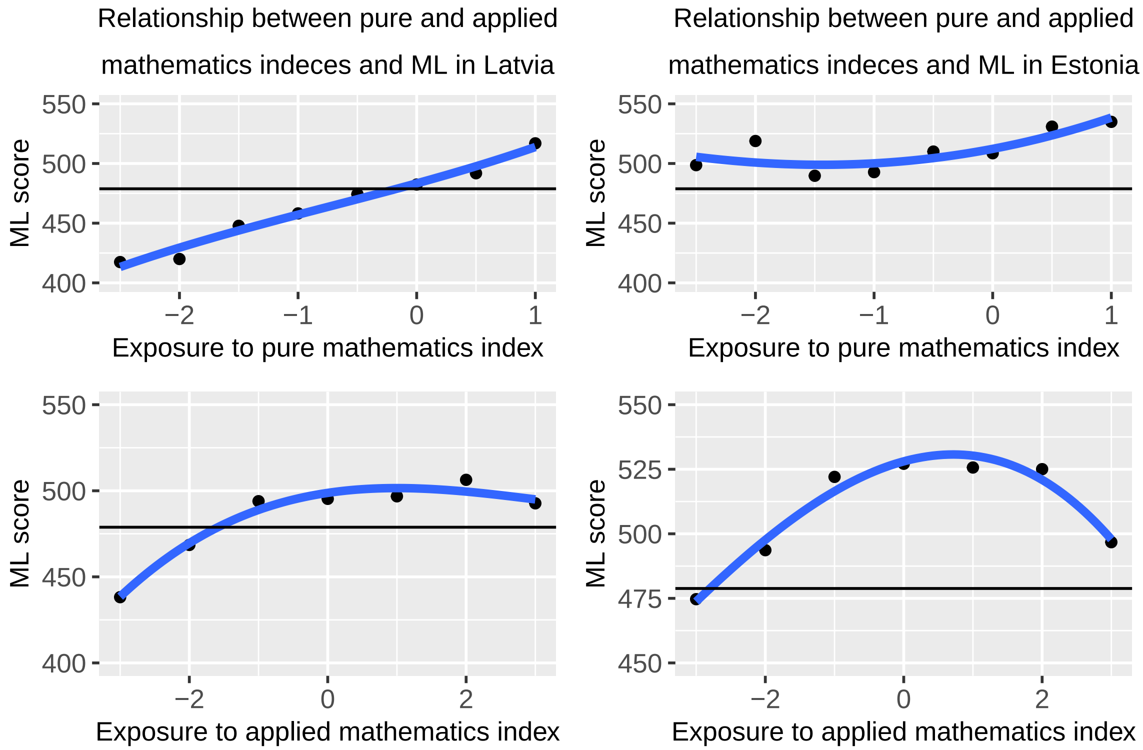

Figure 4: Relationship between pure and applied mathematics indices and ML in Latvia and Estonia

The one unit increase in SES will result in 27-31.5 points increase in ML score, while in Latvia and Estonia the ML score will increase by 20-27 points. Another reason why there are SES differences may be bigger social segregation of students in Lithuania.

The comparison of indices of exposure to pure and applied mathematics support the study hypothesis, however, the impact of exposure to pure mathematics on ML in Latvia and Estonia is much lower. While exposure to pure mathematics estimates in Lithuania are between 23.96-26.7 points, Latvia and Estonia has 17-86-19.95 and 12.55-13.19 accordingly, whereas parameters are statistically significant at $p < 0.001$. The exposure to applied mathematics has a negative relationship with performance in mathematics, however, all but one are statistically insignificant.

The relationship between exposure to pure mathematics and ML score are positive, while in Latvia seem to be almost linear, whereas in Estonia e quadratic (positive). The slopes of exposure to pure mathematics in Latvia and Estonia are flatter than in Lithuania. It can be the a consequence of overall higher results in Latvia and especially Estonia. It was also observed by PISA that in countries with high average ML score exposure to pure mathematics has bigger impact on low performing students, rather than top performers, therefore countries with higher ML average score would have flatter slope on exposure to pure mathematics.

The quadratic (negative) relationship of applied mathematics and ML score was also observed in Latvia and Estonia. Quadratic relationship of applied mathematics in Latvia has little influence on students' ML scores, when in Estonia average ML score start rapidly decreasing after index of exposure to applied mathematics passes value of 0.

7 Conclusions

1. The hypothesis of the study cannot be rejected – exposure to pure mathematics has a positive and consistent relationship with mathematics literacy score on population level in Lithuania. Such relationship is statistically significant, while exposure to applied mathematics has negative (quadratic form) relationship which is usually insignificant. It is supported by the multilevel models for survey data with different model specification: random intercept with fixed slopes and both random intercepts and slopes. Risks from analysis of survey data with two-stage design and two-level weights was controlled using weight scaling procedure and conditional weights. Multilevel modeling with conditional weights resulted in higher intraclass correlation compared to model where weights scaling procedure was used. Hereby estimates from four different multilevel models were compared with each other and with linear regression model as final step of sensitivity analysis. Results from all 5 models support the hypothesis: a one unit increase in index of exposure to pure mathematics will increase student ML score by 18-27 points. The relationship is resilient even after controlling for students' socioeconomic status and grade.

2. The negative relationship of exposure to applied mathematics with mathematics literacy score in Lithuania appear to be "semi-quadratic": it has a positive relationship with mathematics literacy score, but after a certain level of exposure of applied mathematics index is reached e ML starts to decrease. The analysis of Latvia and Estonia shows that index of exposure to applied mathematics follows same pattern, but even greater. This relationship was also confirmed by literature and PISA reports.

3. The hypothesis of the study holds for all sub-areas of mathematics. The estimate of exposure to pure mathematics coefficient vary between 24 and 29 points, meaning that 1 unit increase in index of exposure to pure mathematics will increase student ML score by 24-29 points. Estimate of exposure to applied mathematics coefficient was negative in all cases and usually insignificant. Those results were expected because mathematics literacy score and its sub-scores are correlated.

4. The analysis of other related factors with mathematics area suggest that learning time per week is not statistically significant on population level, while familiarity with mathematics concepts and interest in mathematics have statistically significant positive relationship with mathematics literacy score. It was confirmed using multilevel model and linear regression model.

5. A statistically significant positive relationship of pure mathematics and ML holds in all regions, types of schools and SES groups. However, those results must be treated more carefully, because

the survey design was developed to analyze population level data with possible analysis at school level, not higher levels.

6. The future research can be continued with more detailed analysis of weight scaling techniques as well as school-level cluster analysis in Lithuania in order to evaluate best model and, hereby, unbiased estimates and standard errors (or with minimal bias). It is also possible to include school-level data and analyze relationship between school-level variables and results in any PISA subject literacy score.

8 Appendix

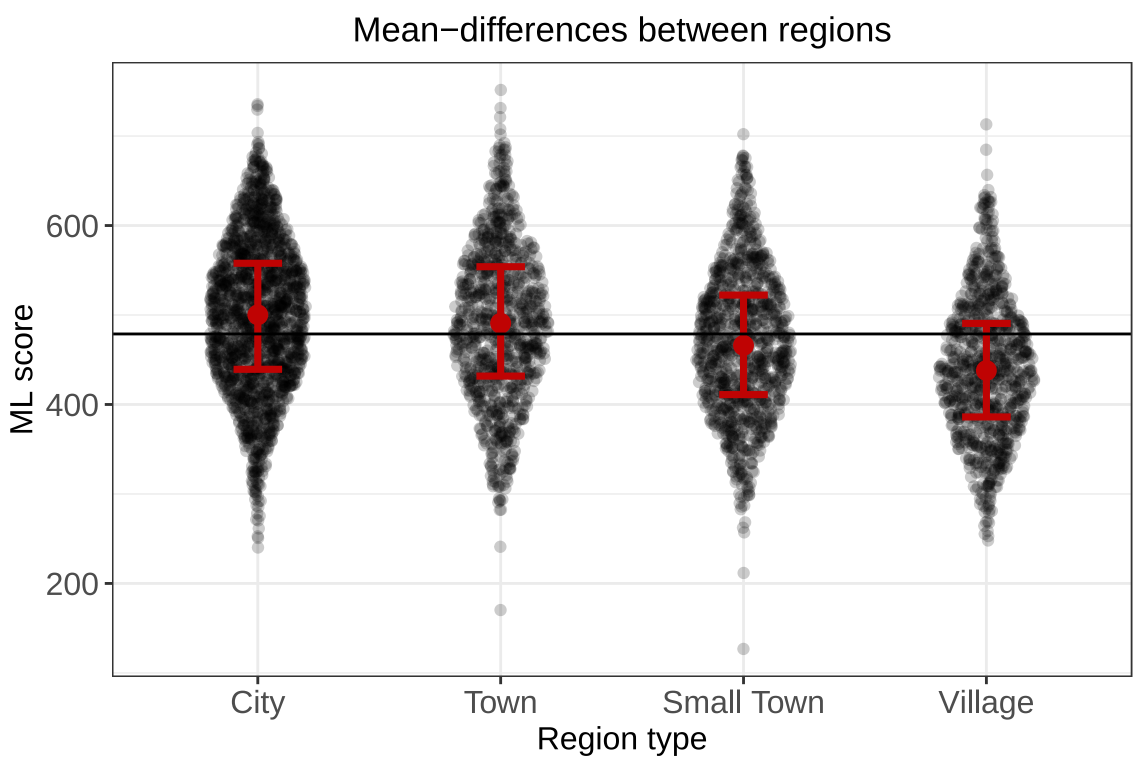

Figure 5: Mean differences of ML scores in Lithuania by region type

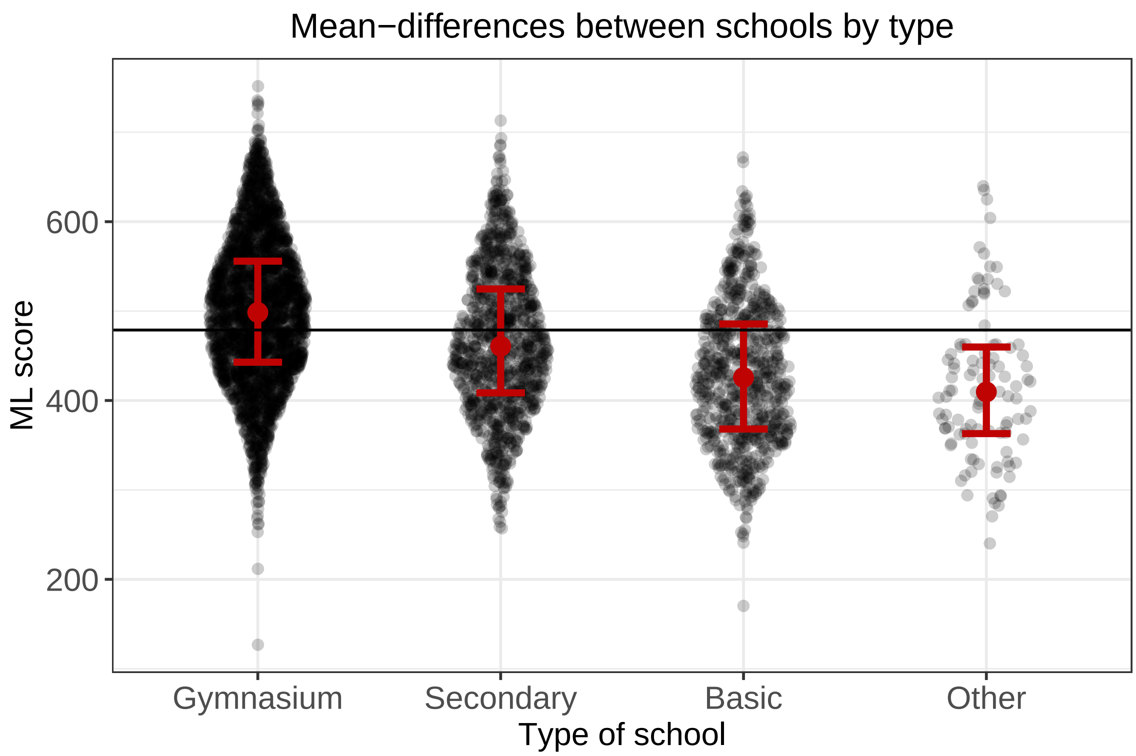

Figure 6: Mean differences of ML scores in Lithuania by school type

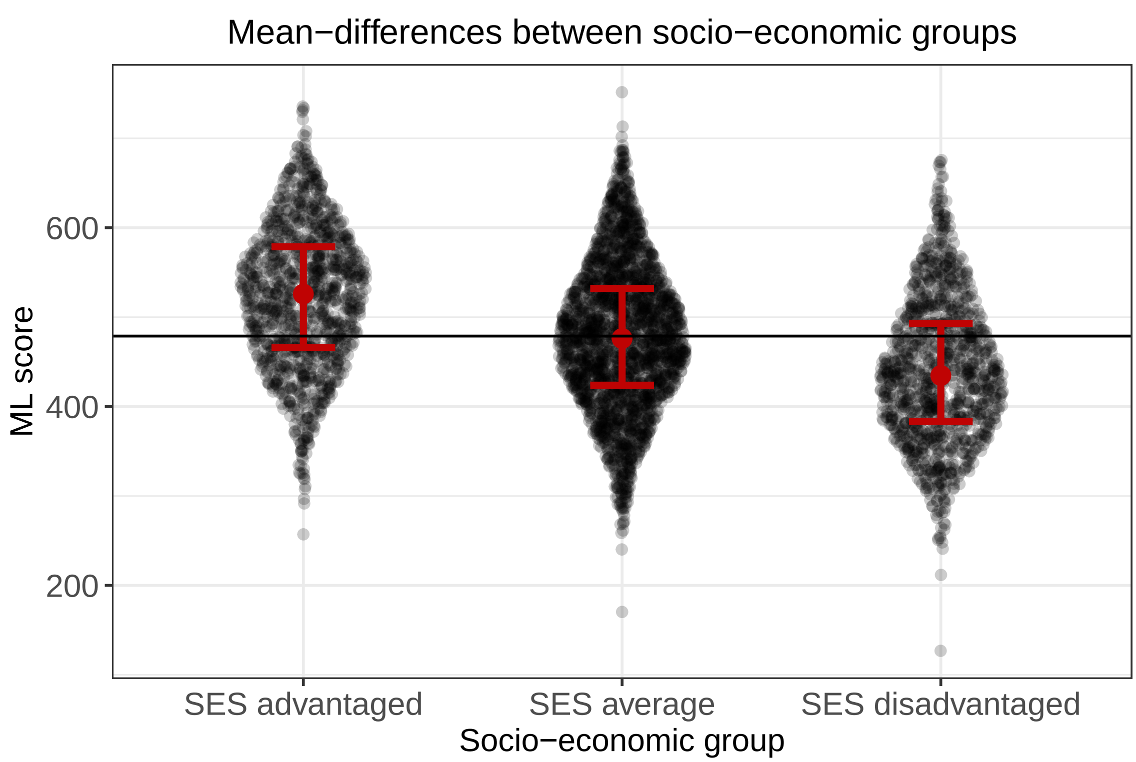

Figure 7: Mean differences of ML scores in Lithuania by socioeconomic groups

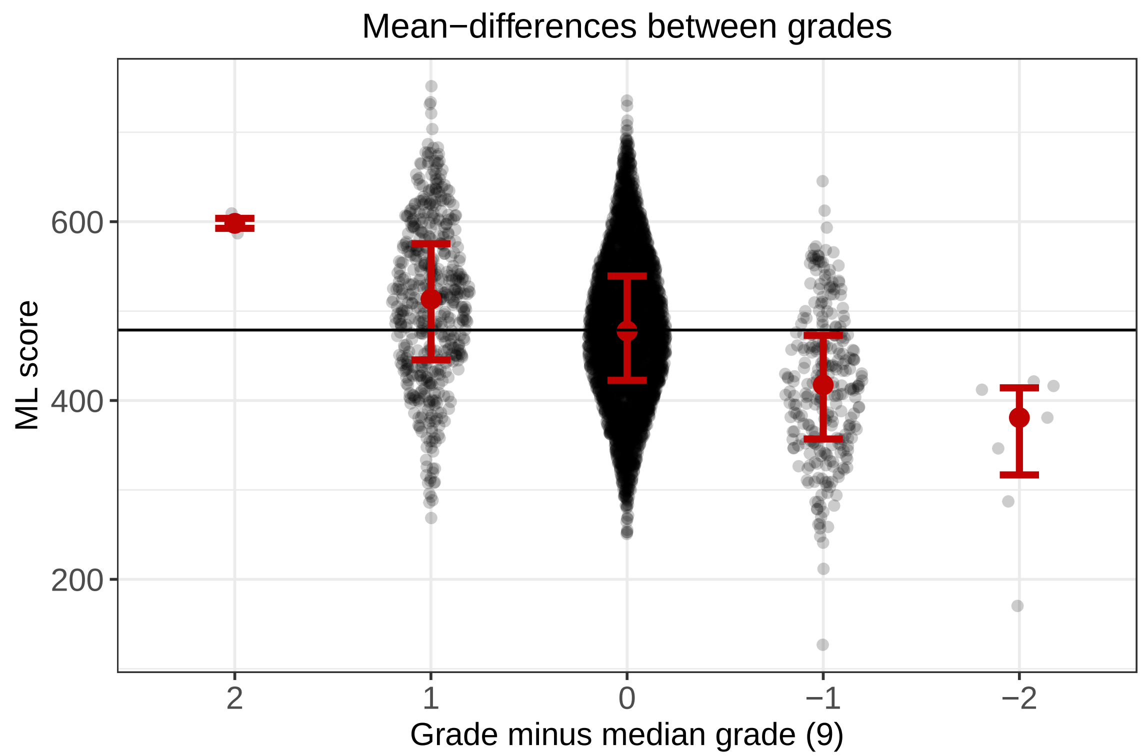

Figure 8: Mean differences of ML scores in Lithuania by student grades

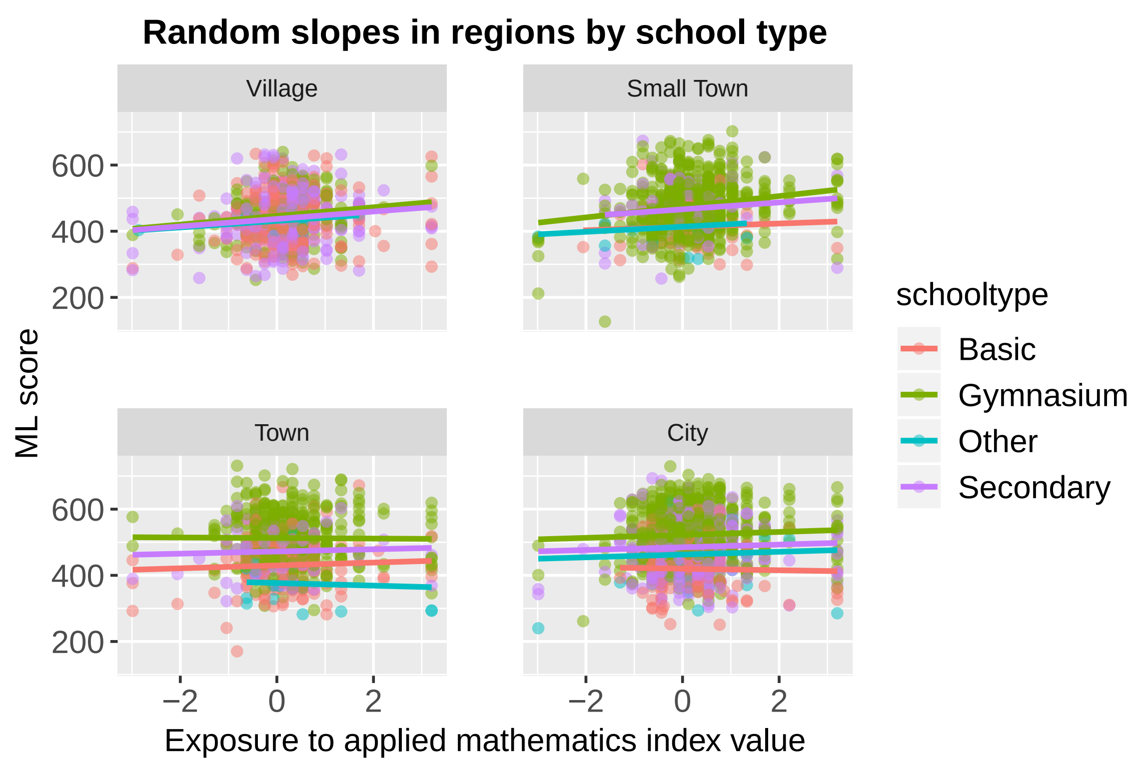

Figure 9: Random slopes for exposure to applied mathematics index in regions for every type of school

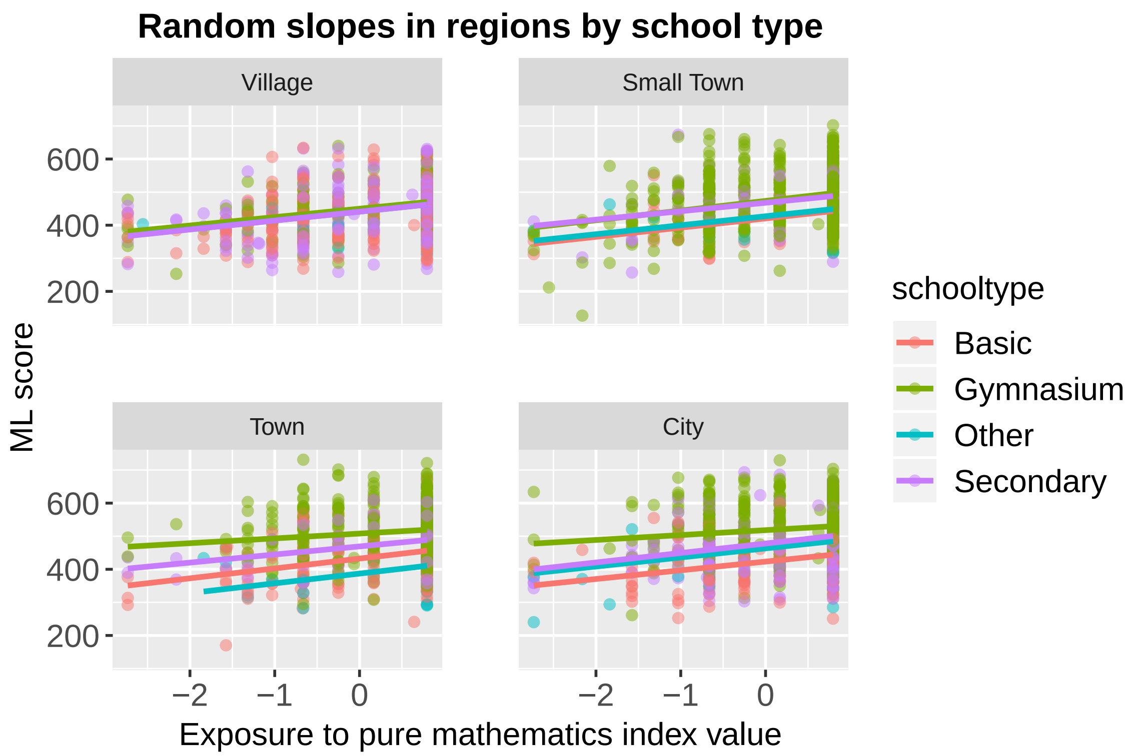

Figure 10: Random slopes for exposure to pure mathematics index in regions for every type of school

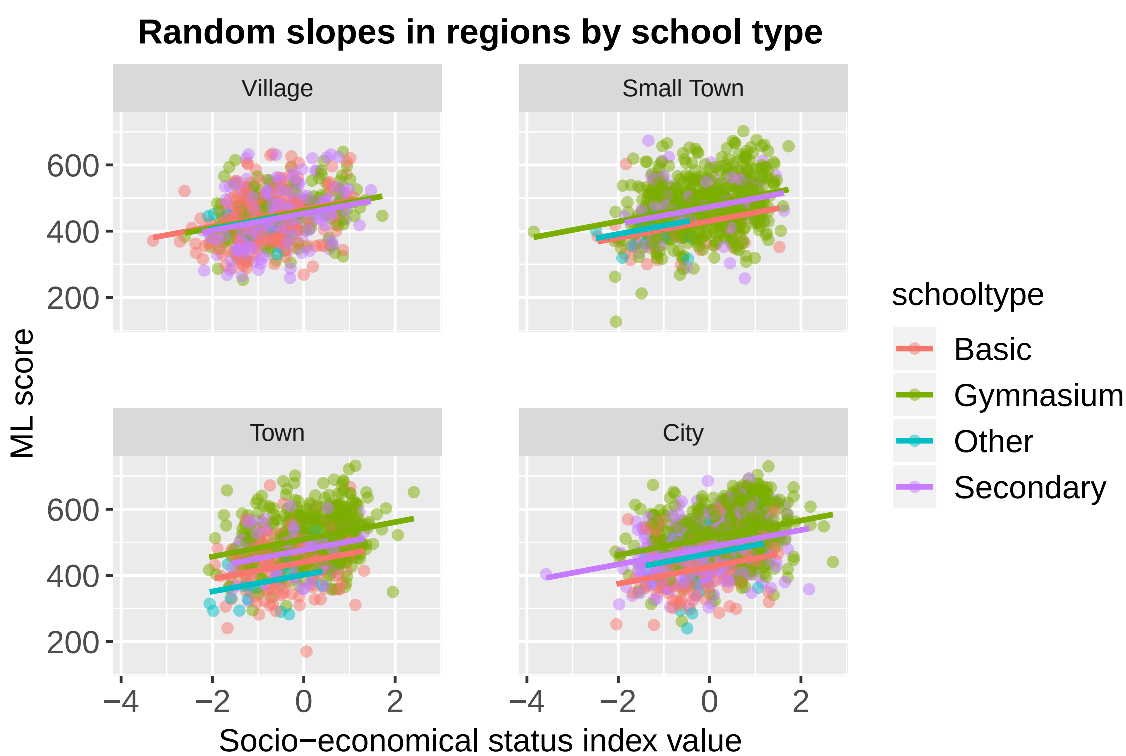

Figure 11: Random slopes for SES index in regions for every type of school

	math	*macc*	*macq*	*macs*	*macu*	*mape*	*mapf*	*mapi*
math	1.000	0.945	0.946	0.918	0.938	0.963	0.945	0.946
macc	0.945	1.000	0.923	0.890	0.916	0.934	0.911	0.920
macq	0.946	0.923	1.000	0.885	0.916	0.940	0.912	0.917
macs	0.918	0.890	0.885	1.000	0.888	0.898	0.912	0.871
macu	0.938	0.916	0.916	0.888	1.000	0.913	0.912	0.928
mape	0.963	0.934	0.940	0.898	0.913	1.000	0.931	0.938
mapf	0.945	0.911	0.912	0.912	0.912	0.931	1.000	0.918
mapi	0.946	0.920	0.917	0.871	0.928	0.938	0.918	1.000

Table 9: Multivariate regression residual correlation matrix

References

OECD and PISA Tests are Damaging Education Worldwide - Academics. The Guardian, Guardian News and Media, May 2014. Retrieved December 13, 2019, from www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics.

R. J. Adams and M. L. Wu. PISA 2000 Technical Report. Non-journal, OECD, Paris, 2002. Retrieved October 11, 2019, from: http://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/33688233.pdf.

D. P. Baker, B. Goesling, and G. K. LeTendre. Socioeconomic status, school quality, and national economic development: A cross-national analysis of the "Heyneman-Loxley Effect" on mathematics and science achievement. *Comparative education review*, 46(3):291–312, 2002.

A. E. Beaton. Implementing the New Design: The NAEP 1983-84 Technical Report. Non-journal, National Assessment of Educational Progress, Princeton, NJ, March 1987. Retrieved December 13, 2019, from: https://eric.ed.gov/?id=ED288887.

R. Breen and J. O. Jonsson. Inequality of opportunity in comparative perspective: Recent research on educational attainment and social mobility. *Annu. Rev. Sociol.*, 31:223–243, 2005.

M. Broer, Y. Bai, and F. Fonseca. *Socioeconomic Inequality and Educational Outcomes: Evidence from Twenty Years of TIMSS*, volume 5. Springer, 2019.

D. Cēdere, I. Jurgena, and V. Targamadze. Interest of Latvian and Lithuanian students in science and mathematics. *Journal of Baltic Science Education*, 17(1):31–42, 2018.

M. M. Chiu and L. Khoo. Effects of resources, inequality, and privilege bias on achievement: Country, school, and student level analyses. *American Educational Research Journal*, 42(4):575–603, 2005.

L. S. Cogan, W. H. Schmidt, and S. Guo. The role that mathematics plays in college-and career-readiness: evidence from PISA. *Journal of Curriculum Studies*, 51(4):530–553, 2019.

Y. Kim. Revisiting classroom practices in East Asian Countries: Examination of within-country variations and effects of classroom instruction. *Teachers College Record*, 120(7):1–42, 2018.

LRT. Putinaitė: kaimo mokyklose negali būti gerų rezultatų, kad ir kokie mokytojai ten bedirbtų, 2019. Retrieved December 14, 2019, from: https://www.lrt.lt/naujienos/lietuvoje/2/1062665/putinaite-kaimo-mokyklose-negali-buti-geru-rezultatu-kad-ir-kokie-mokytojai-ten-bedirbtu.

Y. Lu and D. M. Bolt. Examining the attitude-achievement paradox in PISA using a multilevel multidimensional IRT model for extreme response style. *Large-scale Assessments in Education*, 3(1):2, 2015.

J. Mikk. Economic and educational correlates of TIMSS results. *Socialiniai tyrimai*, (7):83–89, 2006.

I. V. Mullis, M. O. Martin, P. Foy, and M. Hooper. TIMSS 2015 international results in mathematics., 2016.

G. Murauskas and M. Radavičius. Studentų pasiekimų analizė naudojant ranginę logistinę regresiją. *Lietuvos matematikos rinkinys*, 51:408–413, 2010.

B. Narkevičienė and R. Novikienė. Matematikos didaktikos mokslas Lietuvoje: Quo vadis? *Lietuvos matematikų draugijos darbai*, 59(B):67–75, 2018.

Rimas Norvaiša. Why do we teach the mathematics that we do? *Lietuvos matematikos rinkinys*, 60(A):21–26, 2019.

OECD. *PISA 2006: Science Competencies for Tomorrow's World: Volume 1: Analysis*. PISA. OECD Publishing, Paris, 2007. doi: https://doi.org/10.1787/9789264040014-en.

OECD. *PISA Data Analysis Manual: SPSS, Second Edition*. PISA. OECD Publishing, Paris, 2009. doi: https://doi.org/https://doi.org/10.1787/9789264056275-en.

OECD. *PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I)*. PISA. OECD Publishing, Paris, 2010.

OECD. *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. PISA. OECD Publishing, Paris, 2013a. doi: https://doi.org/10.1787/9789264190511-en.

OECD. *PISA 2012 Results: What Students Know and Can Do (Volume I, Revised edition, February 2014): Student Performance in Mathematics, Reading and Science*. PISA. OECD Publishing, Paris, 2013b.

OECD. PISA 2012 Technical Report. Non-journal, OECD, 2014. Retrieved September 13, 2019, from: https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf.

OECD. *PISA 2015 results (Volume I): Excellence and equity in education*. PISA. OECD Publishing, Paris, 2016a.

OECD. *Equations and Inequalities: Making Mathematics Accessible to All*. PISA. OECD Publishing, Paris, 2016b.

OECD. *Education in Lithuania*, volume Reviews of National Policies for Education of *PISA*. OECD Publishing, Paris, 2017.

OECD. *PISA 2018 Results (Volume I): What Students Know and Can Do*. PISA. OECD Publishing, Paris, 2019.

L. B. Perry and A. McConney. Does the SES of the school matter? An examination of socioeconomic status and student achievement using PISA 2003. *Teachers College Record*, 112(4):1137–1162, 2010.

D. Pfeffermann, C. J. Skinner, D. J. Holmes, H. Goldstein, and J. Rasbash. Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society: series B (statistical methodology)*, 60(1):23–40, 1998.

S. Rabe-Hesketh and A. Skrondal. Multilevel modelling of complex survey data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(4):805–827, 2006.

M. Radavičius and P. Samusenko. Palyginamoji pažangumo analizė: panelinės logistinės regresijos modelis. *Lietuvos matematikos rinkinys*, 51:373–378, 2010.

S. W. Raudenbush and A. S. Bryk. *Hierarchical linear models: Applications and data analysis methods*, volume 1. Sage, 2002.

W. H. Schmidt and A. Maier. Opportunity to Learn. In G. Sykes, B. Schneider, and D. N. Plank, editors, *Handbook of Education Policy Research*, pages 541–559. Routledge, 2009.

F. I. Stevens. Applying an opportunity-to-learn conceptual framework to the investigation of the effects of teaching practices via secondary analyses of multiple-case-study summary data. *The journal of Negro education*, 62(3):232–248, 1993.

V. Strauss. Perspective | Expert: How PISA Created an Illusion of Education Quality and Marketed It to the World. The Washington Post, WP Company, December 2019. Retrieved December 13, 2019, from https://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics.

L. M. Thien. Malaysian studentsâ performance in mathematics literacy in PISA from gender and socioeconomic status perspectives. *The Asia-Pacific Education Researcher*, 25(4):657–666, 2016.

D. Tiešytė et al. Matematika mokykloje, matematika universitete. Ar lengva peršokti į naujas roges? *Alfa plius omega*, 12(2):59–66, 2001.

B. D. Wright and M. H. Stone. *Best test design*. Mesa press, Chicago, IL, 1979.

R. Želvys et al. The PISA Phenomenon: The Many Faces of International Student Assessment. *Acta Paedagogica Vilnensia*, (37):9–17, 2016.