

UNIVERZITET U SARAJEVU

PRIRODNO – MATEMATIČKI FAKULTET

ODSJEK ZA FIZIKU

II CIKLUS STUDIJA – FIZIKA U OBRAZOVANJU

**SEKUNDARNA ANALIZA REZULTATA PISA
2018 ISTRAŽIVANJA: IMPLIKACIJE ZA
OBRAZOVANJE IZ PRIRODNIH NAUKA U BOSNI
I HERCEGOVINI**

ZAVRŠNI RAD II CIKLUSA STUDIJA

Mentori: prof. dr. Ali Eryılmaz,
prof. dr. Vanes Mešić

Kandidatkinja:
Semra Sikira

Sarajevo, 2021.

UNIVERSITY OF SARAJEVO

FACULTY OF SCIENCE

DEPARTMENT OF PHYSICS

II CYCLE OF STUDIES – PHYSICS EDUCATION

**SECONDARY ANALYSIS OF PISA 2018
RESULTS: IMPLICATIONS FOR SCIENCE
EDUCATION IN BOSNIA AND HERZEGOVINA**

FINAL THESIS OF THE SECOND CYCLE OF STUDIES

Mentors: Prof. Dr. Ali Eryılmaz,
Prof. Dr. Vanes Mešić

Candidate:
Semra Sikira

Sarajevo, 2021

ABSTRACT

In 2018, 6480 15-year-old students from Bosnia and Herzegovina for the first time ever, took part in the Programme for International Student Assessment (PISA). Students from Bosnia and Herzegovina scored significantly below average in scientific literacy. The aim of this study was to investigate the factors that explain variability in the scientific literacy scores of the students who participated in PISA 2018 in Bosnia and Herzegovina, as to unravel their implications on the science education in Bosnia and Herzegovina.

Variables of interest were determined from the relevant literature review and grouped by the Supply-Use model of schooling. Hierarchical Linear Modeling was used to analyze 22 student-level and 13 school-level factors, and their interactions. The results showed 13 student-level, 5 school-level and 2 interaction effects as significant predictors of scientific literacy:

Student factors: gender, grade repetition, time spent in science classes, immigration status, parents' emotional support, socio-economic and cultural status, discriminating school environment, attitude towards learning, expected occupational status, motivation to master tasks, fear of failure, meaning in life and positive feelings.

School factors: class size, amount of extra-curricular activities, proportion of teachers with Master diploma, shortage of educational material and teacher behavior hindering learning.

Interaction factors: socio-economic and cultural status of the student and the amount of extra-curricular activities at school, immigration status of student and proportion of teachers with Master diploma in school.

Implications of the findings were made for: science teachers, principals of schools, parents of students, and policy makers in Bosnia and Herzegovina.

Keywords: Scientific Literacy, Programme for International Student Assessment (PISA), Bosnia and Herzegovina, Supply-Use Model, Hierarchical Linear Modelling (HLM), student-level factors, school-level factors

SAŽETAK

Po prvi put u historiji, Bosna i Hercegovina je učestvovala u Programu za međunarodnu procjenu učeničkih postignuća, PISA 2018. U ovom istraživanju učestvovalo je 6480 učenika dobi od 15 godina sa prostora cijele Bosne i Hercegovine. Rezultati učenika iz Bosne i Hercegovine su bili naročito poražavajući u domenu prirodnoznastvene pismenosti. Cilj ove studije bio je istražiti faktore koji objašnjavaju varijabilnost u rezultatim naučne pismenosti studenata koji su učestvovali u PISA 2018 istraživanju, u svrhu otkrivanja implikacija na obrazovanje iz prirodnih nauka u Bosni i Hercegovini.

Varijable od interesa utvrđene su iz pregleda relevantne literature i grupisane prema modelu Ponude-Upotrebe školskog učenja. Hijerarhijsko linearno modeliranje korišteno je za analizu 22 učenička faktora i 13 školskih faktora, te njihove interakcije. Rezultati su ukazali na 13 relevantnih učeničkih, 5 školskih faktora i 2 interakcije kao značajne prediktore naučne pismenosti:

Učenički faktori: spol, ponavljanje razreda, vrijeme provedeno na časovima prirodnih nauka, imigracijski status, emocionalna podrška roditelja, socio-ekonomski i kulturološki status, diskriminirajuće školsko okruženje, odnos prema učenju, očekivani status zanimanja, motivacija za savladavanje zadataka, strah od neuspjeha, životna svrha i pozitivna osećanja.

Školski faktori: veličina razreda, količina vannastavnih aktivnosti, udio nastavnika sa master diplomom, nedostatak obrazovnog materijala i ponašanje nastavnika koje ometa učenje.

Interagirajući faktori: socio-ekonomski i kulturni status učenika i količina vannastavnih aktivnosti u školi, te imigracijski status učenika i udio nastavnika sa završenom master školom.

Implikacije koje proizilaze iz ovih pronalazaka odnose se na: nastavnike prirodnih nauka, direktore škola, roditelje učenika i kreatore edukacionih politike u Bosni i Hercegovini.

Ključne riječi: Naučna pismenost, Program za međunarodno ocjenjivanje učenika (PISA), Bosna i Hercegovina, Model ponude-upotrebe, Hijerarhijsko linearno modeliranje (HLM), faktori na nivou učenika, faktori na nivou škole



To my brother, Semir – my biggest academic inspiration.

ACKNOWLEDGMENTS

Finding a good mentor in academia is a challenge big enough, but to be blessed with two great mentors, as were mine, is a strike of luck. I would like to thank Prof. Dr Ali Eryılmaz from Middle East Technical University, Turkey for his interest, shared knowledge, endless patience, and encouragement during my thesis writing. Prof. Dr Vanes Mešić from University of Sarajevo, Bosnia and Herzegovina, has been one of the most influential people in my journey in Physics Education, and I am beyond grateful for his dedication and belief in me. Their critique, suggestions, and recommendations made this thesis possible.

I would also like to deeply thank Ulaş Üstün, a professor from Artvin Çoruh University, Turkey, for his selfless support of my journey with Hierarchical Linear Modelling, both conceptually and pragmatically. Three of them were always there for me; whenever I needed help, to guide me in the right direction. I can only hope to always be surrounded with mentors with such pedagogical, methodological and before anything else, kind approach to their mentees.

I would also like to thank Žaneta Džumhur for her insight and suggestions about my thesis topic. Yasin, Semir, Amina, Murisa and my mother all contributed to my thesis with their discussions, readings, and suggestions. My parents always supported me through my academic pursuit and I hope I have made them proud. Finally, I would like to thank Akın, for being with me and for supporting me emotionally throughout the challenges I faced during my thesis writing.

TABLE OF CONTENTS

ABSTRACT	iii
SAŽETAK.....	iv
ACKNOWLEDGMENTS	vi
LIST OF TABLES	x
LIST OF FIGURES	xi
1. INTRODUCTION.....	1
1.1 Background of the Study	1
1.2 Purpose of the Study.....	4
1.3 Significance of the Study.....	7
2. LITERATURE REVIEW.....	8
2.1 Scientific literacy	8
2.1.1 Scientific literacy according to PISA.....	10
2.2 Studies of Scientific Literacy and School Achievement	13
2.2.1 Student – level factors.....	13
2.2.2 School – level factors.....	14
2.3 General Model of Education.....	16
2.3.1 Supply – Use model of School Learning	17
2.4 Summary of Literature Review	18
3. METHODS	20
3.1 Design of the Study	20
3.2 Population and sample.....	20
3.2.1 Sample weighing in PISA	21
3.3 Variables	21

3.3.1 Explanatory (predictor) variables	23
3.4 Instruments	33
3.4.1 Quality monitoring, Validity and Reliability	34
3.5 Procedures	34
3.5.1 Analysis strategy	35
3.5.2 Assumptions for the Hierarchical Linear Modeling (HLM)	37
4. RESULTS.....	39
4.1 Data Cleaning and Missing Data Analysis	39
4.2 Descriptive Statistics	41
4.3 Inferential Statistics	45
4.3.1 Verification of the HLM assumptions	45
4.3.2 Results for Research Question I (Null model)	48
4.3.3 Results of Research Question II (student-level model)	49
4.3.4 Results of Research Question III (school-level model)	51
4.3.5 Random coefficient model	53
4.3.6 Full model (Interaction model)	58
5. DISCUSSION, CONCLUSIONS, and IMPLICATIONS	62
5.1 Discussion of the Results	62
5.1.1 Limitations of the Study	62
5.1.2 Integration of Findings from Student-level Factors with the Related Literature Review	63
5.1.3 Integration of Findings from School-level Factors with the Related Literature Review	65
5.2 Conclusions	66
5.3 Implications	68
5.3.1 Suggestions for Further Research	69

REFERENCES	70
APPENDICES	79
APPENDIX A.....	79
A.1 Descriptive Statistics	79
A.1.1 Histograms for Level-1 predictors	79
A.1.2 Histograms for Level-2 predictors	82
A.2.3 Descriptive statistics (before Expectation Maximization)	85
APPENDIX B.....	87
B.1 Slope Variability	87

LIST OF TABLES

Table 1.1 The Present Study Model

Table 2.1 Summary description of the seven levels of science proficiency in PISA 2018

Table 4.1 Descriptive statistics for level-1 variables

Table 4.2 Descriptive statistics for level-2 variables

Table 4.3 Bivariate correlations of the level-1, level-2 predictor variables and one (PV1SCIE) criterion variable.

Table 4.5 Results of the Null Model

Table 4.6 Null model for variance components

Table 4.6 Results of the Fixed effects model for level-1 variables

Table 4.7: Results of the school level (level-2) model

Table 4.8: Results of the random effects model

Table 4.9 Results of the Final Model

Table 5.1 The Present Study Model (Results)

Table A. 1 Descriptive statistics (level-1 data) before Expectation-Maximization missing data replacement

Table A. 2 Descriptive statistics (level-2 data) before Expectation-Maximization missing data replacement

LIST OF FIGURES

Figure 2.1 Multilevel Supply-Use model of student learning adapted from (Brühwiler & Blatchford, 2011) based on (Helmke, 2012, p. 71) and (Baumert et al., 2004, p. 149).

Figure 4.1 Scatterplot of the regression residuals to the predicted value

Figure 4.2 Normal Q-Q plot of the residuals



1. INTRODUCTION

1.1 Background of the Study

“Education is the most powerful weapon which you can use to change the world” –

Nelson Mandela

Humans have been learning and developing their knowledge of the world and natural phenomena ever since the dawn of time. Transferring that knowledge from generation to generation, world developed, countries formed and there came about a need for a structured approach to learning. Roles of students and teachers became formal. Schools formed. Formal education was born.

Quality of a country's education has been shown to have far reaching consequences on all aspects of its society (Barrett et al., 2006). An organization which recognizes the importance of education in a country's development and prosperity is the Organization for Economic Co-operation and Development (OECD). It was founded in 1961 with a goal to promote sustainable economic growth, support trade, partake in policy formation and help developing countries.

An OECD project which educators, governments and public worldwide are highly aware of is the Programme for the International Assessment of Students (PISA). 79 countries with a 31 million population of 15-year-olds participated in the 2018 cycle of PISA. The student sample included approximately 710 000 students (OECD, 2016a).

PISA is a collaborative effort of OECD and its partner countries in "assessing the extent to which 15-year-old students near the end of their compulsory education have acquired the knowledge and skills essential for full participation in modern societies." (OECD, 2016a, p. 3) It is held every three years from its conception in 1997 and it primarily measures the levels of: reading, mathematical and scientific literacy. PISA produces a large amount of data that is put at the disposal of researchers. In the context of this master thesis, it is valuable to mention that "secondary analysis" is the most published type of article related to PISA, taking 61.8% of all published articles in English language by the year 2015 (Hopfenbeck et al., 2018).

In 2018, for the first time in its history, 6489 15-year-old students in Bosnia and Herzegovina took part in the PISA research. PISA 2018 scientific literacy in Bosnia and Herzegovina of 398 was well below OECD standard (489) and one of the lowest in the Balkan region. Definition of scientific literacy in PISA 2018 is given below:

„Scientific literacy: The ability to engage with science-related issues, and with the ideas of science, as a reflective citizen. A scientifically literate person is willing to engage in reasoned discourse about science and technology, which requires the competencies to explain phenomena scientifically, evaluate and design scientific enquiry, and interpret data and evidence scientifically.” (OECD, 2016a, p. 16)

The average score of students of Bosnia and Herzegovina does not even reach the Level 2 of OECD’s science proficiency benchmark value. In addition, it is far below the Level 4, which is the OECD average for PISA 2018 scientific literacy.

“Level 2 students are able to draw on everyday content knowledge and basic procedural knowledge to identify an appropriate scientific explanation, interpret data and identify the question being addressed in a simple experimental design. They can use basic or everyday scientific knowledge to identify a valid conclusion from a simple data set. Level 2 students demonstrate basic epistemic knowledge by being able to identify questions that can be investigated scientifically.” (OECD, 2019f, p.113)

It would be interesting to explore how certain general student factors such as: gender, age and grade repetition affect scientific literacy (Finn, 1989; Gottfredson et al., 1994; Marks, 2006; Sun et al., 2012; Üstün et al., 2020). Individual characteristic of students, such as: motivation, self-confidence, attitudes and learning strategy have also been found to influence science achievement (Husremović & Đapo, 2013; Kremer & Walberg, 1981; Mattern & Schau, 2002; Osborne et al., 2003). Another important factor is the learning environment. Learning environment of the student consists of their family, peers, and media according to Helmke (2012), whereas factors that could determine it are: socio-economic status, immigration status, emotional support of parents, which

have also been found to have an influence on overall student achievement and particularly in science. (Hsu, 2007; Mashile, 2001; Sirin, 2005)

Among the school-level variables, good school climate for which both teachers and school classroom context are responsible was also found to have an effect on PISA performance (OECD, 2009a). Classroom context and namely class size has been investigated in the context of school achievement, with somewhat contradictory results, making it an important factor to explore as well (Dynarski et al., 2013; Juan & Visser, 2017; Woessmann & West, 2006). Other school related factors such as number of computers connected to the internet in Turkey (Üstün et al., 2020) and condition of school infrastructure in South Africa (Juan & Visser, 2017) were found to be significant predictors of science achievement.

Grouping student- and school level variables based on a theoretical model would be useful as to put the present study in the context of general models of school learning. The Supply-Use model was found to be the closest fit to the aim of this study (Baumert et al., 2004; Brühwiler & Blatchford, 2011; Helmke & Lenske, 2013; Seidel, 2015). In the model, the *supply* is related to the school and its characteristics such as: teacher characteristics and classroom context as well as learning processes, whereas the *use* is related to: learning environment, individual student characteristics and individual learning processes.

Except the formal report done by the Agency for preschool, primary school and secondary school education (APOS0) (Džumhur, 2019) on PISA 2018 results and some benchmark analyses related to mathematic and reading literacy (Husremović & Đapo, 2013), PISA's scientific literacy in Bosnia and Herzegovina has not been deeply researched. The goal of the present study is to fill that gap in the literature.

1.2 Purpose of the Study

PISA 2018 average scientific literacy score for students in Bosnia and Herzegovina (398) was well below OECD standard (489) and one of the lowest in the Balkan region. Most of the students from Bosnia and Herzegovina do not even achieve the proficiency Level 2 – a benchmark in scientific literacy and they are even farther from the proficiency Level 4 which is the OECD PISA 2018 scientific literacy average. Since PISA measures *skills essential for full participation in modern societies*, the scores of students from Bosnia and Herzegovina are deeply concerning. Deeper research of the PISA 2018 results was primarily focused on its main domain, which was reading literacy.

The aim of this study was to explore which student and school factors, as well as their interactions have significant relationships with Bosnia and Herzegovina students' scientific literacy scores as measured in PISA 2018. Variables divided into two levels and grouped based on the theoretical Supply-Use model were analyzed for significant between-school, within-school variance, and their interaction on the scientific literacy scores with Hierarchical Linear Modeling.

Table 1.1 The Present Study Model

Student – level variables (level-1)
<i>General student characteristics:</i>
Student gender
Grade compared to modal grade in country
Age
Grade repetition
Minutes of Student's time spend in science classes per week
<i>Learning environment</i>
Perception of competitiveness at school
Perception of cooperation at school
Discriminating school climate
Parents' emotional support perceived by student
Index of economic, social, and cultural status
Immigration status
Student's experience of belonging to school
Student's experience of being bullied
<i>Individual student characteristics:</i>
Competitiveness
Motivation to master tasks
General fear of failure
Eudaemonia: meaning in life
Subjective well-being: Positive feelings
Resilience
Mastery goal orientation
Student's expected occupational status
Attitude towards school: learning activities
School – level variables (level-2)
<i>General school context:</i>
Proportion of available computers that are connected to the Internet
Quantity of creative extra-curricular activities
Shortage of educational material
Shortage of educational staff
Student behavior hindering learning
School principal's view on teachers' multicultural and egalitarian beliefs
<i>Teacher characteristics</i>
Total number of all teachers at school
Proportion of all teachers fully certified
Proportion of all teachers with Bachelor degree
Proportion of all teachers with Master degree
Proportion of all teachers with PhD
Teacher behavior hindering learning
<i>Classroom context:</i>
Class Size

Note: The Supply – Use model on school achievement was used primarily for the grouping of the student – level and the school – level variables.

List of research questions the present study aimed to answer:

Research Question I: *What are between-school and within-school variances in PISA2018 scientific literacy scores of 15-year-old students in Bosnia and Herzegovina?*

Research Question II: *Which student-level variables predict scientific literacy scores of 15-year-old students as measured in PISA2018 in Bosnia and Herzegovina?*

Research Questions III: *Which school-level variables predict scientific literacy scores of 15-year-old students as measured in PISA2018 in Bosnia and Herzegovina?*

Research Questions IV: *Which combination of school-level and student-level variables predicts scientific literacy scores of 15-year-old students measured in PISA2018 in Bosnia and Herzegovina?*

Research Questions V: *Are there any significant interactions between student-level and school-level variables that predict scientific literacy scores of 15-year-old students as measured in PISA2018 in Bosnia and Herzegovina have significant interactions?*

1.3 Significance of the Study

Considering the significance of PISA and its effects on educational policies worldwide, deeper analysis of PISA 2018 results for Bosnia and Herzegovina could be of great value for science teachers, educational policy makers and other researchers. The fact that students from Bosnia and Herzegovina scored particularly low in scientific literacy posed many important questions about possible actions that could increase the quality of science education in Bosnia and Herzegovina. Understanding the challenges of science education can provide new opportunities in the context of scientific literacy (McFarlane, 2013).

The present study also aims to use as much of PISA 2018 data for Bosnia and Herzegovina that can be related to scientific literacy, considering that the main domain of PISA 2018 was reading literacy.

Present study, to the knowledge of the author, is the first of its kind for a deeper analysis of scientific literacy of the PISA data for Bosnia and Herzegovina. Even though these types of studies with similar methodology were done in other countries, each country and its education systems are unique. Therefore, present study could give more insight for the specific educational context in Bosnia and Herzegovina. It could also serve as a benchmark study compared to the upcoming PISA cycles, which would hopefully be done as well in Bosnia and Herzegovina.

2. LITERATURE REVIEW

Literature review starts with the historic development of the scientific literacy concept and then goes on to the specifics of PISA's scientific literacy framework. In the second part, studies that concern separate student and school level factors which affect scientific literacy and school achievement overall are presented divided into two levels. In the last part, general models of education are presented with focus on the so-called Supply-Use model, which takes into consideration the hierarchical nature of educational system.

2.1 Scientific literacy

At the end of the 18th century Johann Lorenz Boeckmann in his book "Science about nature" argues the necessity of concisely studying natural sciences in schools. He argues that studying science helps us in: expanding our horizons, perfecting trade, and arts, getting rid of childish fears and superstitions, and even conceiving God and its creation.

As the industrialization became more prominent and the trade grew, the need for educated individuals in the field of natural sciences increased. Therefore, schooling also experienced the movement towards the fulfillment of those needs. Europe certainly took its time in accepting the value of science education and its implementation to the already existing schooling systems. However, on the other side of the Atlantic, a more pragmatic theory of education was given more attention. The socio-economic view of this theory was arguably the most important aspect, considering the industrialization in which a value of a man was equated to his professional success. (Kircher et al., 2009). John Dewey (1923) enriched this approach by bringing in the values which laid the foundations for the modern educational system in North America. According to him, the methods of natural sciences represent the most powerful approach of acquiring knowledge. Therefore, he concurs that the main goal of Physics education, for example, is to teach students to use the methods of Physics itself, rather than to just learn facts about Physics.

The world after the Second World War put more emphasis on technological development and national security. "The goals of science teaching for general education purposes within this new environment came to be called scientific literacy" (DeBoer, 2000, p. 586). One of the first, if

not *the* first time “Scientific Literacy” was used in a publication, was in the 1958 by Paul Hurd in the publication *Scientific literacy: Its meaning for Americas’ Schools* (Hurd, 1958).

Definition of scientific literacy is not universally agreed upon (Laugksch, 2000). However, Showalter (1974) (as cited in Rubba & Andersen, 1978) tried to integrate 15 years’ worth of literature in his scientific literacy definition bellow:

“Definition of scientific literacy consisting of seven dimensions:

- I. The scientifically literate person understands the nature of scientific knowledge.
- II. The scientifically literate person accurately applies appropriate science concepts, principles, laws, and theories in interacting with his universe.
- III. The scientifically literate person uses processes of science in solving problems, making decisions, and furthering his own understanding of the universe.
- IV. The scientifically literate person interacts with the various aspects of his universe in a way that is consistent with the values that underlie science.
- V. The scientifically literate person understands and appreciates the joint enterprises of science and technology and the interrelationship of these with each and with other aspects of society.
- VI. The scientifically literate person has developed a richer, more satisfying, more exciting view of the universe as a result of his science education and continues to extend this education throughout his life.
- VII. The scientifically literate person has developed numerous manipulative skills as-sociated with science and technology.” (Rubba & Andersen, 1978, p. 450)

It is obvious that society appreciated the economic aspect of scientific literacy even from the early 19th century, however, how does that translate to the modern world of the 21st century? An answer to this, an organization called the OECD is trying to give for the last 20 years.

The Organization for Economic Co-operation and Development (OECD) was formally founded on the 30th of September 1961. Their predecessor, The Organization for European Economic Co-operation (OEEC) was established in 1948. The goal of OEEC was to rebuild Europe after the Second World War, thereby ensuring lasting peace. By encouraging reconstruction and co-operation, OECD promotes sustainable economic growth, supports trade, partakes in policy formation and helps developing countries (OECD, 2020).

2.1.1 Scientific literacy according to PISA

A project which educators, governments and public worldwide are highly aware of is the OECDs' Programme for the International Assessment of Students (PISA). PISA is a collaborative effort of OECD and its partner countries in "assessing the extent to which 15-year-old students near the end of their compulsory education have acquired the knowledge and skills essential for full participation in modern societies." (OECD, 2016, p. 11) It is held every three years from its conception in 1997 and it measures the levels of: reading, mathematical and scientific literacy. In each cycle one of the major domains is extensively tested. In 2003 and 2012 mathematics was the core domain, while in 2006 and 2015 science was. In the last 2018th cycle of PISA, together with the ones from 2000 and 2009, reading was the focus of the study (OECD, 2016a). For this study it is particularly important to describe how PISA conceptualizes the construct of scientific literacy.

According to PISA 2018:

“Scientific literacy: The ability to engage with science-related issues, and with the ideas of science, as a reflective citizen. A scientifically literate person is willing to engage in reasoned discourse about science and technology, which requires the competencies to explain phenomena scientifically, evaluate and design scientific enquiry, and interpret data and evidence scientifically. (OECD, 2016a, p. 16)

PISA assesses students' performance in science through questions related to:

- **Contexts:** These include personal, local/national, and global issues, both current and historical, that demand some understanding of science and technology.
- **Knowledge:** This is the understanding of the major facts, concepts and explanatory theories that form the basis of scientific knowledge. Such knowledge includes knowledge of both the natural world and technological artefacts (content knowledge), knowledge of how such ideas are produced (procedural knowledge), and an understanding of the underlying rationale for these procedures and the justification for their use (epistemic knowledge).
- **Competencies:** These are the ability to explain phenomena scientifically, evaluate and design scientific enquiry, and interpret data and evidence scientifically.” (OECD, 2016a, p. 16)”

An important concept of PISA is also their so-called *proficiency levels*.

“In PISA, the reporting scales are called “proficiency scales” rather than “performance scales” because they describe what students at given levels of proficiency typically know and can do, rather than how individuals who were tested actually performed on a single test administration. This emphasis reflects the primary goal of PISA, which is to report general population-level results rather than the results for individual students.” (Ward, 2018, Chapter 15, p. 1) (Table 2.1).

Particularly, “Level 2 in science is an important benchmark for student performance: it represents the level of achievement, on the PISA scale, at which students begin to demonstrate the science competences that will enable them to engage in reasoned discourse about science and technology. (OECD, 2016b, p.113)” Proficiency levels are there to ensure easier comparability of PISA data across different countries and educational systems.

Table 2.1 Summary description of the seven levels of science proficiency in PISA 2018

Level	Lower score limit	Cumulative percentage*	Characteristics of tasks
6	708	0.8%	At Level 6, students can draw on a range of interrelated scientific ideas and concepts from the physical, life, and earth and space sciences and use content, procedural and epistemic knowledge in order to offer explanatory hypotheses of novel scientific phenomena, events and processes or to make predictions. In interpreting data and evidence, they are able to discriminate between relevant and irrelevant information and can draw on knowledge external to the normal school curriculum. They can distinguish between arguments that are based on scientific evidence and theory and those based on other considerations. Level 6 students can evaluate competing designs of complex experiments, field studies or simulations and justify their choices
5	633	6.8%	At Level 5, students can use abstract scientific ideas or concepts to explain unfamiliar and more complex phenomena, events and processes involving multiple causal links. They are able to apply more sophisticated epistemic knowledge to evaluate alternative experimental designs and justify their choices, and use theoretical knowledge to interpret information or make predictions. Level 5 students can evaluate ways of exploring a given question scientifically and identify limitations in interpretations of data sets, including sources and the effects of uncertainty in scientific data.

4	559	24.9%	At Level 4, students can use more complex or more abstract content knowledge, which is either provided or recalled, to construct explanations of more complex or less familiar events and processes. They can conduct experiments involving two or more independent variables in a constrained context. They are able to justify an experimental design by drawing on elements of procedural and epistemic knowledge. Level 4 students can interpret data drawn from a moderately complex data set or less familiar context, draw appropriate conclusions that go beyond the data and provide justifications for their choices.
3	484	52.3%	At Level 3, students can draw upon moderately complex content knowledge to identify or construct explanations of familiar phenomena. In less familiar or more complex situations, they can construct explanations with relevant cueing or support. They can draw on elements of procedural or epistemic knowledge to carry out a simple experiment in a constrained context. Level 3 students are able to distinguish between scientific and non-scientific issues and identify the evidence supporting a scientific claim.
2	410	78.0%	At Level 2, students are able to draw on everyday content knowledge and basic procedural knowledge to identify an appropriate scientific explanation, interpret data and identify the question being addressed in a simple experimental design. They can use basic or everyday scientific knowledge to identify a valid conclusion from a simple data set. Level 2 students demonstrate basic epistemic knowledge by being able to identify questions that can be investigated scientifically
1a	335	94.1%	At Level 1a, students are able to use basic or everyday content and procedural knowledge to recognise or identify explanations of simple scientific phenomena. With support, they can undertake structured scientific enquiries with no more than two variables. They are able to identify simple causal or correlational relationships and interpret graphical and visual data that require a low level of cognitive demand. Level 1a students can select the best scientific explanation for given data in familiar personal, local and global contexts.
1b	261	99.3%	At Level 1b, students can use basic or everyday scientific knowledge to recognise aspects of familiar or simple phenomena. They are able to identify simple patterns in data, recognise basic scientific terms and follow explicit instructions to carry out a scientific procedure.

Adapted from (OECD, 2019f, p.113)

Note:* Percentage of students able to perform tasks at each level or above (OECD average)

2.2 Studies of Scientific Literacy and School Achievement

PISA measures scientific literacy on a student level, however it also measures many other types of variables concerning school climate, affective characteristics and general information about students and school on both student and school level. This section deals with student and school specific factors that affect science achievement and scientific literacy. Factors which are found to be good predictors of school achievement overall, are also included.

2.2.1 Student – level factors

There is contradicting proof about the effect of gender on scientific literacy. For PISA 2018 females outperformed males by 2 points on the OECD's average on scientific literacy scores (Džumhur, 2019). However some other studies of PISA scientific literacy, show that gender gap does exist in favor of males (Sun et al., 2012; Üstün et al., 2020). In addition, studies of gender gap relating to science interest show that in early childhood and most of the primary school there is no difference in science interest among genders. That however starts to change in high school, when males end up significantly more interested in science than females until the end of high school. (Baram-Tsabari & Yarden, 2010)

Even though PISA measures scientific literacy of 15-year-olds specifically, effect of grade is proved to be significant, since students can even be in different levels of education, like primary school and secondary school (Džumhur, 2019; Üstün et al., 2020). It has been shown that even if students are of the same age, the ones in higher grades perform better than those in lower grades (Marks, 2006). Repetition of grade, in theory, would allow students to better learn what they were not able to and catch up to their peers. However, students that repeat a grade are more likely to be stigmatized, leading to worse behavior and attitudes towards school, even increasing their risk of leaving school altogether (Finn, 1989; Gottfredson et al., 1994; Jacob & Lefgren, 2004).

Learning environment, as a system consisting of: family, peers and media (Helmke & Lenske, 2013), have been shown as an influential factor in science achievement (Kidron & Lindsay, 2014). Socio-economic status (SES) is one of the major themes in educational research (Sirin, 2005) and it even receives dedicated attention in the PISA design having its own category as an ESCS index. SES of a student is likely the most important factor correlated with the performance

in PISA, being responsible for about 20% student score variation (OECD, 2009a). In addition to other family dimensions like style of parenting, involvement of parents and parents' attitude (i.e., emotional support) are also shown to be significantly related to students' science achievement (Ho, 2010; Mashile, 2001). Immigration status has also been shown to be a factor in school achievement. For example, in both Canada and Hong Kong's PISA2000 results, immigration status (being non-native) was a significant negative predictor of mathematical literacy scores (Hsu, 2007).

There is a body of research that shows that there are significant correlations between student characteristics such as SES, educational level of parents, attitudes etc. (Atwater et al., 1995; Ceylan & Abaci, 2013; Luu & Freeman, 2011; Özer & Anil, 2011; Watters & Ginns, 2000). It has been found that parents' attitudes towards science have a positive correlation with science achievement throughout many countries (Simpson & Oliver, 1990). However even this factor does not mediate the SES-achievement relationship (Perera, 2014).

According to Ozel (2013) factors such as: curriculum, learning, environment, demographic and affective factors are associated with academic achievement. Affective factors have an even bigger support in literature as being a predictor of academic achievement. (Ceylan & Berberoglu, 2007; Gungor et al., 2007; Mettas et al., 2006; Nasr & Soltani, 2011; Rennie & Punch, 1991; Singh et al., 2002; Uguroglu & Walberg, 1979; Usak et al., 2009; Willson et al., 2000). Self-confidence, motivation and attitudes towards school are shown to have strong relationships with performance as well (Husremović & Đapo, 2013).

Body of literature that is making a case for the importance of affective factors in science education is significant as well. The relationship between science achievement and affective factors can give a deeper understanding to the teaching and learning process (Mattern & Schau, 2002). Among affective factors, motivation has been deeply researched, finding a positive relationship with science achievement (Kremer & Walberg, 1981). It has also been shown that that relationship gets stronger with age (Uguroglu & Walberg, 1979).

2.2.2 School – level factors

In the first three cycles of PISA, it has been shown that the between-school variability in students' scores accounts for 30% of all variability in students' scores. While percentage in the northern European countries is below 20%, in countries like Austria, Belgium and Germany more

than 50% of variability in student performance is found between schools. The strongest unique factor would be the social intake. This effect called "the school composition effect" tries to explain how higher performing students likely come from more advantaged schools, as opposed to the less advantaged schools. Higher performance is also found in private schools. However, when taking into consideration factors such as socio-economic status of the student and demographics this factor ceases to be statistically significant. Even though there is a moderate relationship between performance and school resources, it can yet again be explained by the socio-economic variable. Student performance tends to be negatively influenced by a small degree when there is grouping by ability in schools. The school climate created by teachers and students, tends to be an important factor which affects school performance, correlating with students' background characteristics such as SES again (OECD, 2009a).

Studies about effect of class size on school achievement are contradictory. Some show the advantage of smaller class sizes (Dynarski et al., 2013), some are neutral (Woessmann & West, 2006) and there is also a case of bigger class sizes like they are typically in Asian countries that overperform in PISA (Džumhur, 2019). Condition of school buildings has been found to be significant indicator of science achievement, as shown with South African students in TIMSS 2012 (Juan & Visser, 2017). In addition, the number of computers that are connected to the internet has been found to be a significant variable for scientific literacy (Üstün et al., 2020). Lastly, quality of teachers education was found to affects student academic achievement as well (Hsu, 2007).

2.3 General Model of Education

In a search for factors which would be the possible predictors of science achievement measured by PISA 2018, a need for a theoretical model or framework that explains students' academic achievement arose. This section gives an overview of different theoretical models of school learning (Huitt & McIlrath, 1995). Starting from Carroll (1963), then Bloom (1971) all the way to Huitt (1995) and then Helmke (2012) many researchers proposed different models of school learning.

John Carroll (1963) theorized that time spent learning is the most important factor in school learning. According to him, time was dependent upon factors such as: student's IQ, student's prerequisite knowledge and the quality of instruction received by the student. He did not consider factors such as: influence of family, peers, and society to be significant in predicting school learning.

On the other hand, Bloom observed that student's aptitude for learning academic material (i.e., IQ) was one of the best predictors of school achievement. He also demonstrated in his research that if students are given sufficient time to learn with quality instruction, they can almost all preform way above average (Bloom et al., 1971). According to Bloom, attitudes, and quality of instruction play an important role in school achievement. 25% of between-students variance in science achievement could be attributed to how students feel towards what they are studying, their school environment and their self-concept. Another 25% of variance could be attributed to quality of instruction, whereas 50% would be attributed to the cognitive characteristic. (Simpson & Oliver, 1990)

Proctor (1984) put emphasis on teacher traits that facilitate learning. His model acknowledges how school social climate (i.e., norms, the attitudes, and the values of schools) can make an impact on the achievement test scores as well as their correlation with student's achievement level.

Huitt's (1995) main contribution was the importance of context (i.e., family, home, school, and peers) to the school achievement. In his framework he shows how variables related to context, input (students and teachers' contribution), classroom processes (what happens in the classroom) and output (i.e., achievement test as a measure of learning) have an intertwined relationship.

Utilization-of-Learning-Opportunities model (Helmke, 2012), also referred to as the Supply–Use model (Brühwiler & Blatchford, 2011) was found to be the best approximation of a theoretical model for the present study. Not only was it used in some other PISA related research (Lenzen et al., 2004), but it also fits the most of the data already offered in the PISA 2018 dataset for Bosnia and Herzegovina.

2.3.1 Supply – Use model of School Learning

According to Helmke (2012), learning environment secures learners with opportunities (supply) for learning. However, whether the acquiring of the knowledge (process of learning, i.e., use) happens depends on utilizing of said opportunities. The responsibility of utilizing the learning opportunities lies on the learner itself, indicating how learning is an interaction between a learner and a context. This framework is named the Supply-Use model of school learning, and it is presented in the Figure 2.1.

The *supply* level of the model considers the qualities of the learning opportunities. Beginning with the characteristics of an educational system, school and subject contextualize the opportunities to learn. Inside of that context, further characteristics of teachers and classroom context are considered to influence classroom processes. Professional competencies of teachers such as: their knowledge, motivation, and beliefs, but also their general characteristics such as age and gender influence instructional quality. However, classroom context factors such as: heterogeneity, class size and overall achievement affect also contribute to the classroom processes in an interaction with teacher characteristics.

The *use* level is determined by the individuals learning processes occurring during the instruction and individual characteristics of the student that are influenced by the learning environment. Even though learning environment (family, peers, media) is not related to the factors that the individual can necessarily control, it still influences individual student characteristic. Factors such as: cognitive, motivational, emotional, and self-related cognitions are determinators of the individual student characteristics.

Lastly, the *outcomes* level is related to the individual learning activities related to the individual students' characteristics. Learning outcomes can be cognitive as measured with

standardized test scores, but can also be motivation-affective and meta-cognitive.

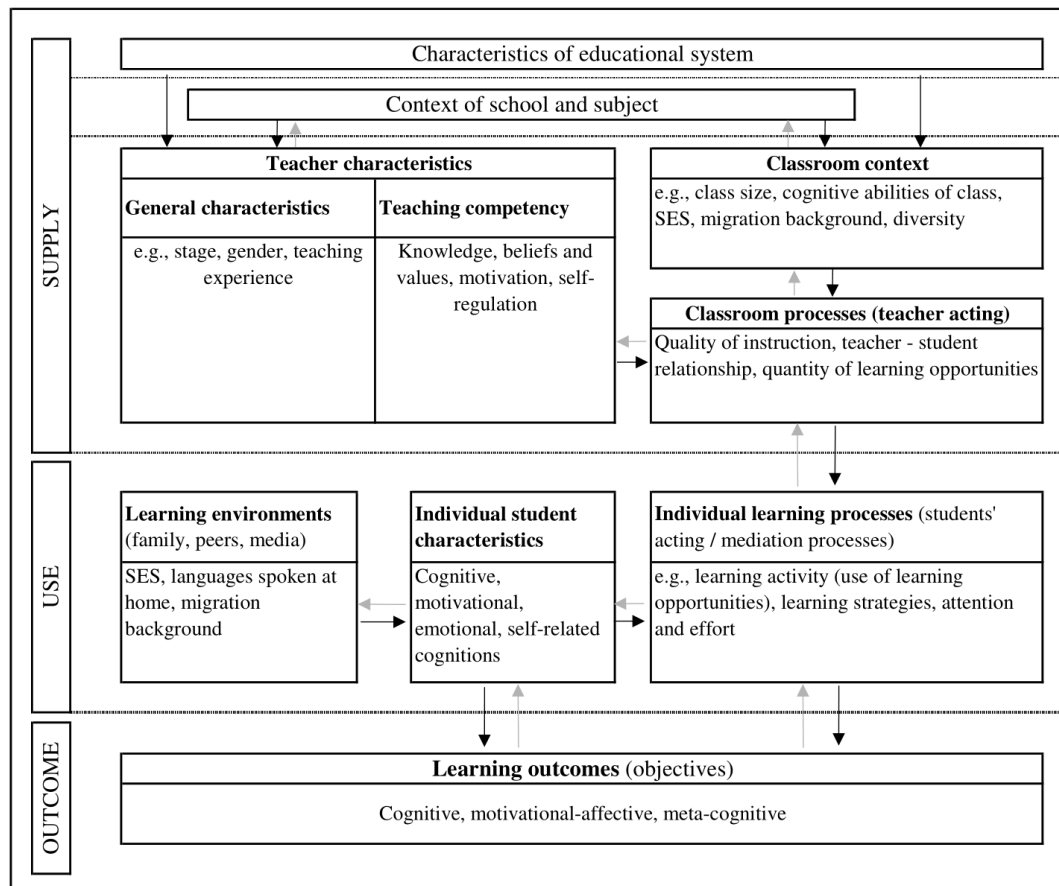


Figure 2.1 Multilevel Supply-Use model of student learning adapted from (Brühwiler & Blatchford, 2011) based on (Helmke, 2012, p. 71) and (Baumert et al., 2004, p. 149).

2.4 Summary of Literature Review

PISA's scientific literacy is one of many possible measures of science achievement. Based on this review of the literature, it can be concluded that there are many mutually dependent student and school factors that influence science achievement. Out of the student-level factors, literature indicates that general student characteristics such as: gender, age and grade repetition can be meaningful to explore (Finn, 1989; Gottfredson et al., 1994; Marks, 2006; Sun et al., 2012; Üstün et al., 2020).

Learning environment of the student consists from their family, peers, and media according to Helmke (2012), whereas factors that could determine it are: socio-economic status, immigration

status, emotional support of parents, which have also been found to have an influence on student achievement and science achievement (Hsu, 2007; Mashile, 2001; Sirin, 2005).

Individual student characteristics of students, such as: motivation, self-confidence, attitudes and learning strategy have also been found to influence science achievement (Husremović & Đapo, 2013; Kremer & Walberg, 1981; Mattern & Schau, 2002; Osborne et al., 2003).

Continuing onto school-level variables, themselves less researched than student level variables, indicate another level of approach to explaining variability in academic achievement and science scores. Earlier research shows that socio-economic status predicts higher performance, as well as good school climate for which both teachers and school administration are responsible. (OECD, 2009a) Effect of class size is ambiguous but still an important factor (Brühwiler & Blatchford, 2011; Glass & Smith, 1979; Juan & Visser, 2017; Woessmann & West, 2006). Lastly, number of computers connected to the internet have been found to be significant as well for scientific literacy (Üstün et al., 2020).

Some general models and frameworks of school learning are presented as well, showing how learning environment characteristics and students emotional processes can be a significant predictor of achievement. Namely the Supply-Use model and its use in literature showed reasoning behind multilevel design in educational research (Brühwiler & Blatchford, 2011; Helmke & Lenske, 2013; Seidel, 2015). The *supply* is related to school and its characteristics such as: teacher characteristics and classroom context as well as learning processes, whereas the *use* is related to: learning environment, individual student characteristics and individual learning processes. Supply-Use models also serves as an example of a way to group various student and school factors that explain cognitive outcomes.

3. METHODS

This chapter begins with research methodology, continuing onto description of the population and sample. Following is the variables section, going into more detail about predictor and criterion variables with their descriptions, minimum, maxima and reliability coefficients where applicable. Next subsection concerns the instrument with which needed data was gathered, including quality monitoring, validity, and reliability of the PISA2018 study. Final subsection deals with procedures; it provides a detailed explanation of the analysis strategy used in this study.

3.1 Design of the Study

Organizations such as OECD, through large – scale research such as PISA, offer huge amounts of secondary data for countries to be able to do analyses that could have a meaningful impact on their educational system. Quantitative study with correlational or associational methodology was used in the present study. Correlational research attempts to determine how and to what extent are the two or more variables related (Gall et al., 1996). It is the most common non-experimental research design that arises in educational research (Karasar, 2012).

Two datasets provided by the OECD with variables concerning student and school characteristics are used in a Multilevel Modelling (MLM), precisely Hierarchical Linear Modeling (HLM) to decompose factors which affect scientific literacy in Bosnia and Herzegovina. Reason for using HLM is the nested structure of data (i.e., students nested in schools), for which single – level analyses are not fitting. Predictor variables were divided into two levels: student – level and school level, whereas criterion variable was scientific literacy in the form of plausible values.

3.2 Population and sample

In a period between 2nd of April and 25th of May 2018, 6480 15-year-old students from 213 primary and secondary schools in Bosnia and Herzegovina participated in the PISA 2018 research. At the time of the collection of the population information in Bosnia and Herzegovina, there were 863 (1817 if regional schools are included) primary and 311 secondary schools with 282947

students in primary and 124148 students in secondary education. Approximately 8% of these students were 15 years old at the time of research. The sample consisting of 6864 students from 213 schools throughout the country was reduced to 6480 (3148 girls and 3332 boys) due to exclusion according to PISA standards and it represents more than 82% of the whole national 15-year old population (Džumhur, 2019).

Sample includes both governmental and private schools as well as primary and secondary schools of all types and from all cantons of the Federation of Bosnia and Herzegovina, regions from Republic Srpska and Brčko district from rural and urban areas. Private schools with education implemented in foreign languages were excluded as well as special education schools. Less than 3% of schools in BiH were excluded from sampling. In each school 46 15-year-old students were sampled and if the school has less than 46 students, all 15-year-old students were sampled.

3.2.1 Sample weighing in PISA

Even though student and school samples in PISA2018 were acquired through stratified random sampling, their probability of selection varies. Proportional contribution of each student or school to the overall population estimate is not the same. This is just one of the reasons for which PISA suggest using weights while doing analyses on their databases. To avoid biased parameter estimates, it is suggested to use the final student weights (W_FSTUWT) which can be found in the Student Questionnaire Dataset (OECD, 2019e) (OECD, 2019d). Other types of weights, as well as their use will be discussed in the Section 5.1.1.

3.3 Variables

This section gives an overview and a deeper insight into the types of variables used in this study. PISA's description of what the variable measures and what are the possible values as well as the possible minimum and maximums and the scale of measurement are given for each variable. For the variables which used Weighted Likelihood Estimation (WLE) the Cronbach alpha (α) specifically for Bosnia and Herzegovina is given as well. Plausible Values (PV) are explained as to further give understanding of the criterion variables.

"PISA assessment provides three main types of outcomes:

- Basic indicators that provide a profile of the knowledge and skills of students
- Indicators derived from the questionnaires that show how such skills relate to various demographic, social, economic, and educational variables.
- Indicators on trends that show changes in outcomes and their distributions, and in relationships between student-, school- and system-level background variables and outcomes." (OECD, 2016a)

Many variables measure latent constructs (like student's motivation or socio – economic status) and are therefore formed as indices. Student – level predictor variables as well as criterion variables are taken from the Student Questionnaire and the school – level predictor variables are taken from the School Questionnaire. There are four types of variables in PISA2018.

First one is simple questionnaire indices which are constructed through arithmetical transformation or recoding of one or more items. Examples of those are: i) for student – level: GRADE, AGE, IMMIG, SMINS, REPEAT, and for school – level: CLSIZE, RATCMP2, TOTAT, PROATCE, PROAT5AB, PROAT5AM, PROAT6, CREATIV.

Second type are derived variables based on the Item Response Theory (IRT) scaling, resulting in the WLE as scores, which are presented with their respective scale reliabilities (Cronbach's Alpha coefficient) for each of the participating countries. As the second type, there are the dispositional variables and school – focused variables. "Dispositional variables are the personality-based context in which students approach learning. They are the result of a lifetime of socialization from parents, teachers, coaches, and one's cultural surroundings, and they capture how behavior is energized over time. School-focused variables, or how students view and approach school, are influenced by both students' disposition and the particular situation in which they find themselves." (OECD, 2019a) For student – level examples of the dispositional variables are: ATTLNACT, COMPETE, WORKMAST, GFOFAIL, RESILIENCE, MASTGOAL, PERCOMP, PERCOOP, EMOSUPS, EUDMO, SWBP, DISCRIM and example of the school – focused variables are: BELONG and BEINGBULLIED. For school – level examples of dispositional variables are: STAFFSHORT, EDUSHORT, STUBEHA, TEACHBEA. Third type is the composite variable that is ESCS, which is calculated from IRT variables.

Fourth type are the Plausible Values (PV's) which are given for the scientific literacy scores. Rasch model (OECD, 2009b) is used to compute Plausible Values. When doing statistical

analyses with Plausible Values it is necessary to do separate analyses with each of the PV's, calculating the average of those analyses at the end. HLM7 software, which was used for this study, has a feature that allows such calculations.

“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 & Wu, 2002)

3.3.1 Explanatory (predictor) variables

3.3.1.1 Student level predictor variables

1. Student (Standardized) Gender (GENDERRCD):

Minimum: 1; Maximum: 2; Scale of Measurement: Nominal

Students were asked to choose one of the two offered genders: (1) female or (2) male. This variable was recoded from its original form (ST004D01T) into GENDERRCD, with following values: (0) female, (1) male.

2. Grade compared to modal grade in country (GRADE):

Minimum: -3; Maximum: 1; Scale of Measurement: Scale

Students were asked to choose the grade they are in. Afterwards the relative grade index (GRADE) was recalculated as to represent whether the student is in the country's modal grade (value of 0), or the one relative to it, i.e., below or above. The meanings of the unique values of GRADE are as follows: Primary school 7th grade (“-3”), Primary school 8th grade (“-2”), Primary school 9th grade (“-1”), Secondary school 1st grade (“0”), Secondary school 2nd grade (“1”).

3. Age (AGE):

Minimum: 14; Maximum: 16; Scale of Measurement: Scale

Student's age was calculated as the difference between the year and month of testing and the year and month of the student's birth.

4. Index Immigration status (IMMIG):

Minimum: 1; Maximum: 3; Scale of Measurement: Ordinal

Students were asked in what country were they and their parents born. Based on their answer the index of immigrant background (IMMIG) was calculated with the following categories: “*native students (those students who had at least one parent born in the country), (2) second-generation students (those born in the country of assessment but whose parent(s) were born in another country) and (3) first-generation students (those students born outside the country of assessment and whose parents were also born in another country).*” (OECD, 2019a) This variable was recoded from its original form (IMMIG) into IMMIGRCD, with following values: (0) native, (1) second-generation, (2) first-generation.

5. Grade Repetition (REPEAT):

Minimum: 0; Maximum: 1; Scale of Measurement: Nominal

Students were asked if they repeated a grade at any of the following levels: ISCED1, ISCED2 or ISCED3. If the student never repeated a grade the value of “0” was assigned to the REPEAT variable, whereas if the student answered that they repeated at least one of the grades, value of REPEAT was assigned “1”.

6. Student's expected occupational status (SEI) (BSMJ):

Minimum: 0; Maximum: 100; Scale of Measurement: Ordinal

“*In PISA 2018, students were asked to answer a question (ST114) about “what kind of job [they] expect to have when [they] are about 30 years old”. Answers to this open-ended question were recoded into two indices related to career expectations. OCOD3 represents the four-digit ISCO codes (ILO, 2007). These ISCO codes were then mapped to the international socio-economic index of occupational status (ISEI) (Ganzeboom & Treiman, 2003) in variable BSMJ. Higher scores on this variable indicate higher levels of a student's expected occupational status.*” (OECD, 2019a)

7. Learning time (minutes per week) - <science> (SMINS):

Minimum: 0; Maximum: unconstrained; Scale of Measurement: Scale

Learning time in science per week was computed by multiplying the number of minutes on average in the science classes with the number of class periods per week.

8. Index of economic, social, and cultural status (ESCS):

Minimum: - 4.79; Maximum: 3.42; Scale of Measurement: Scale; $\alpha = 0.756$

ESCS is an index of economic, social, and cultural status and it is based on education, occupational status, and income of the students' parents. It is calculated as a composite score of: highest parental occupation (HISEI), parental education (PAREDINT), and home possessions (HOMEPOS), including books in the home. More information about the computation of this index can be found in the PISA 2018 technical report 16th chapter. For the scope of this master thesis ESCS was taken as the sole representative of the socio – economic status.

9. Parents' emotional support perceived by student (WLE) (EMOSUPS):

Minimum: 3; Maximum: 12; Scale of Measurement: Scale; $\alpha = 0.894$

Students were asked to rate perceived emotional support from their parents on a four-point Likert scale (“Strongly disagree”, “Disagree”, “Agree”, “Strongly agree”) with the following three statements: “My parents support my educational efforts and achievements.”, “My parents support me when I am facing difficulties at school.”, “My parents encourage me to be confident.”. The values of the EMOSUPS variable were transformed according to the WLE.

10. Perception of competitiveness at school (WLE) (PERCOMP):

Minimum: 4; Maximum: 16; Scale of Measurement: Scale; $\alpha = 0.846$

Students were asked to rate perceived climate of competition in their school on a four-point Likert scale (“Not at all true”, “Slightly true”, “Very true”, “Extremely true”) with the following four statements: “Students seem to value competition.”, “It seems that students are competing with each other”, “Students seem to share the feeling that competing with

each other is important.”, “Students feel that they are being compared with others.” The values of the PERCOMP variable were transformed according to the WLE.

11. Perception of cooperation at school (WLE) (PERCOOP):

Minimum: 4; Maximum: 16; Scale of Measurement: Scale; $\alpha = 0.914$

Students were asked to rate perceived climate of cooperation in their school on a four-point Likert scale (“Not at all true”, “Slightly true”, “Very true”, “Extremely true”) with the following four statements: “Students seem to value cooperation.”, “It seems that students are cooperating with each other”, “Students seem to share the feeling that cooperating with each other is important.”, “Students feel that they are encouraged to cooperate with others.” The values of the PERCOOP variable were transformed according to the WLE.

12. Attitude towards school: learning activities (WLE) (ATTLNACT):

Minimum: 3; Maximum: 12; Scale of Measurement: Scale; $\alpha = 0.846$

Students’ attitude towards school was measured on a reverse-coded four-point Likert scale (“Strongly agree”, “Agree”, “Disagree”, “Strongly disagree”) with the following three statements: “Trying hard at school will help me get a good job.”, “Trying hard at school will help me get into good <college>.”, “Trying hard at school is important.”. The values of the PERCOMP variable were transformed according to the WLE.

13. Competitiveness (WLE) (COMPETE):

Minimum: 3; Maximum: 12; Scale of Measurement: Scale; $\alpha = 0.799$

Students’ competitiveness achievement motive was measured on a four-point Likert scale (“Strongly disagree”, “Disagree”, “Agree”, “Strongly agree”) with the following three statements: “I enjoy working in situations involving competition with others.”, “It is important for me perform better than other people on a task.”, “I try harder when I’m in competition with other people.”. The values of the COMPETE variable were transformed according to the WLE.

14. Work mastery (WLE) (WORKMAST):

Minimum: 4; Maximum: 16; Scale of Measurement: Scale; $\alpha = 0.785$

Students' working motive and mastery achievement motive was measured on a four-point Likert scale ("Strongly disagree", "Disagree", "Agree", "Strongly agree") with the following four statements: "I find satisfaction in working as hard as I can.", "Once I start a task, I persist until it is finished.", "Part of the enjoyment I get from doing things is when I improve on my past performance.", "If I am not good at something, I would rather keep struggling to master it than move on to something I may be good at.". The values of the WORKMAST variable were transformed according to the WLE.

15. General fear of failure (WLE) (GFOFAIL):

Minimum: 3; Maximum: 12; Scale of Measurement: Scale; $\alpha = 0.829$

Students' general fear of failure was measured on a four-point Likert scale ("Strongly disagree", "Disagree", "Agree", "Strongly agree") with the following three statements: "When I am failing, I worry about what others think of me.", "When I am failing, I am afraid that I might not have enough talent.", "When I am failing, this makes me doubt my plans for the future." The values of the GFOFAIL variable were transformed according to the WLE.

16. Eudaemonia: meaning in life (WLE) (EUDMO):

Minimum: 3; Maximum: 12; Scale of Measurement: Scale; $\alpha = 0.894$

Students' sense of meaning and purpose in life (i.e., eudemonia) was measured on a four-point Likert scale ("Strongly disagree", "Disagree", "Agree", "Strongly agree") with the following three statements: "My life has clear meaning or purpose.", "I have discovered a satisfactory meaning in life.", "I have a clear sense of what gives meaning to my life.". The values of the EUDMO variable were transformed according to the WLE.

17. Subjective well-being: Positive affect (WLE) (SWBP):

Minimum: 3; Maximum: 12; Scale of Measurement: Scale; $\alpha = 0.859$

Students' positive affect was measured on a four-point Likert scale ("Never", "Rarely", "Sometimes", "Always") with the following three statements: "Joyful", "Cheerful", "Happy". The values of the SWBP variable were transformed according to the WLE.

18. Resilience (WLE) (RESILIENCE):

Minimum: 5; Maximum: 20; Scale of Measurement: Scale; $\alpha = 0.829$

Students' resilience was measured on a four-point Likert scale ("Strongly disagree", "Disagree", "Agree", "Strongly agree") with the following three statements: "I usually manage one way or another.", "I feel proud that I have accomplished things.", "I feel that I can handle many things at a time.", "My belief in myself gets me through hard times.", "When I'm in a difficult situation, I can usually find my way out of it.". The values of the RESILIENCE variable were transformed according to the WLE.

19. Mastery goal orientation (WLE) (MASTGOAL):

Minimum: 3; Maximum: 15; Scale of Measurement: Scale; $\alpha = 0.897$

Students' mastery-approach orientation of achievement goals was measured on a five-point Likert scale ("Not at all true of me", "Slightly true of me", "Moderately true of me", "Very true of me", "Extremely true of me") with the following three statements: "My goal is to learn as much as possible.", "My goal is to completely master the material presented in my classes.", "My goal is to understand the content of my classes as thoroughly as possible.". The values of the MASTGOAL variable were transformed according to the WLE.

20. Discriminating school climate (WLE) (DISCRIM):

Minimum: 4; Maximum: 12; Scale of Measurement: Scale; $\alpha = 0.868$

Students were asked whether their teachers treat student from all cultural groups with equal respect, therefore measuring the absence of stereotypes, prejudice, and discrimination. Through a four-point Likert scale ("To none or almost none of them", "To some of them", "To most of them", "To all or almost all of them") with the following four statements: "They have misconceptions about the history of some cultural groups.", "They say negative things about people of some cultural groups.", "They blame people of some cultural groups for problems faced by.", "They have lower academic expectations for students of some cultural groups." multicultural school climate was measured. The values of the DISCRIM variable were transformed according to the WLE.

21. Subjective well-being: Sense of belonging to school (WLE) (BELONG):

Minimum: 6; Maximum: 24; Scale of Measurement: Scale; $\alpha = 0.783$

Students' sense of belonging to school was measured on a four-point Likert scale ("Strongly disagree", "Disagree", "Agree", "Strongly agree") with the following six statements: "I feel like an outsider (or left out of things) at school.", "I make friends easily at school.", "I feel like I belong at school", "I feel awkward and out of place in my school.", "Other students seem to like me.", "I feel lonely at school.". The values of the BELONG variable were reverse coded so that higher WLEs correspond to higher level of sense of belonging.

22. Student's experience of being bullied (WLE) (BEINGBULLIED):

Minimum: 6; Maximum: 24; Scale of Measurement: Scale; $\alpha = 0.815$

Students' experience of being bullied was measured on a four-point Likert scale ("Never or almost never", "A few times a year", "A few times a month", "Once a week or more") with the following three statements: "Other students left me out of things on purpose.", "Other students made fun of me.", "I was threatened by other students.". The values of the BEINGBULLIED variable were transformed according to the WLE.

3.3.1.2 School level predictor variables

1. Proportion of available computers that are connected to the Internet (RATCMP2):

Minimum: 0; Maximum: 1; Scale of Measurement: Scale

RATCMP2 was calculated as the ratio of the number of computers available to 15-year-olds for educational purposes to the number of those computers that were connected to the internet, as reported by the school principal.

2. Total number of all teachers at school (TOTAT):

Minimum: 11; Maximum: 136; Scale of Measurement: Scale

Total number of all teachers at school was reported by the school principles.

3. Index proportion of all teachers fully certified (PROATCE):

Minimum: 0; Maximum: 1; Scale of Measurement: Scale

Index proportion of all teachers fully certified was calculated as the ratio of those teachers as reported by school principals to the total number of all teachers at school.

4. Index proportion of all teachers ISCED LEVEL 5A Bachelor (PROAT5AB):

Minimum: 0; Maximum: 1; Scale of Measurement: Scale

Index proportion of all teachers with an ISCED 5A bachelor qualification was calculated as the ratio of those teachers as reported by school principals to the total number of all teachers at school.

5. Index proportion of all teachers ISCED LEVEL 5A Master (PROAT5AM):

Minimum: 0; Maximum: 1; Scale of Measurement: Scale

Index proportion of all teachers with an ISCED 5A master qualification was calculated as the ratio of those teachers as reported by school principals to the total number of all teachers at school.

6. Index proportion of all teachers ISCED LEVEL 6 (PROAT6):

Minimum: 0; Maximum: 1; Scale of Measurement: Scale

Index proportion of all teachers with an ISCED 6 qualification was calculated as the ratio of those teachers as reported by school principals to the total number of all teachers at school.

7. Class Size (CLSIZE):

Minimum: 13; Maximum: 53; Scale of Measurement: Scale

The average class size is reported by the school principles. It is chosen from one of the following nine categories: 13 = “15 students or fewer”, 18 = “16 – 20 students”, 23 = “21 – 25 students”, 28 = “26 – 30 students”, 33 = “31 – 35 students”, 38 = “36 – 40 students”, 43 = “41 – 45 students”, 48 = “46 – 50 students”, 53 = “more than 50 students”.

8. Creative extra-curricular activities (Sum) (CREACTIV):

Minimum: 0; Maximum: 3; Scale of Measurement: Ordinal

The index of extra – curricular activities offered in the school was calculated as the total number of the following activities that occurred at the school: i) band, orchestra, or choir, ii) school play or school musical, and iii) art club or art activities, as reported by the school principals.

9. Shortage of educational material (WLE) (EDUSHORT):

Minimum: 4; Maximum: 16; Scale of Measurement: Scale; $\alpha = 0.842$

Shortage of educational material was measured from the school's principles on which factors hinder school's capacity to provide instruction. EDUSHORT was measured on a four-point Likert scale ("Not at all", "Very little", "To some extent", "A lot") with the following four statements: "A lack of educational material (e.g., textbooks, IT equipment, library or laboratory material).", "Inadequate or poor quality educational material (e.g. textbooks, IT equipment, library or laboratory material).", "A lack of physical infrastructure (e.g. building, grounds, heating/cooling, lighting and acoustic systems).", "Inadequate or poor quality physical infrastructure (e.g. building, grounds, heating/cooling, lighting and acoustic systems)". The values of the EDUSHORT variable were transformed according to the WLE.

10. Shortage of educational staff (WLE) (STAFFSHORT):

Minimum: 4; Maximum: 16; Scale of Measurement: Scale; $\alpha = 0.629$

Shortage of educational material was measured from the school's principal on which factors hinder school's capacity to provide instruction. STAFFSHORT was measured on a four-point Likert scale ("Not at all", "Very little", "To some extent", "A lot") with the following four statements: "A lack of teaching staff.", "Inadequate or poorly qualified teaching staff.", "Inadequate or poorly qualified teaching staff.", "Inadequate or poorly qualified assisting staff.". The values of the STAFFSHORT variable were transformed according to the WLE.

11. Student behavior hindering learning (WLE) (STUBEHA):

Minimum: 6; Maximum: 24; Scale of Measurement: Scale; $\alpha = 0.916$

Student behavior hindering learning was measured from the school principal's perception of student behavior that might influence the provision of instruction at school. STUBEHA was measured on a four-point Likert scale ("Not at all", "Very little", "To some extent", "A lot") with the following six statements: "Student truancy.", "Students skipping classes.", "Students lacking respect for teachers.", "Student use of alcohol or illegal drugs.", "Students intimidating or bullying other students.", "Students not being attentive.". The values of the STUBEHA variable were transformed according to the WLE.

12. Teacher behavior hindering learning (WLE) (TEACHBEH):

Minimum: 5; Maximum: 20; Scale of Measurement: Scale; $\alpha = 0.830$

Teacher behavior hindering learning was measured from the school principal's perception of teacher behavior that might influence the provision of instruction at school. TEACHBEH was measured on a four-point Likert scale ("Not at all", "Very little", "To some extent", "A lot") with the following five statements: "Teachers not meeting individual students' needs.", "Teacher absenteeism.", "Staff resisting change.", "Teachers being too strict with students.", "Teachers not being well prepared for classes." The values of the TEACHBEHA variable were transformed according to the WLE.

13. School principal's view on teachers' multicultural and egalitarian beliefs (WLE) (SCMCEG):

Minimum: 4; Maximum: 16; Scale of Measurement: Scale; $\alpha = 0.890$

School principal's view on teachers' multicultural and egalitarian beliefs was measured on a four-point Likert scale ("Shared among none or almost none of them", "Shared among some of them", "Shared among many of them", "Shared among all or almost all of them") with the following four statements: "It is important for students to learn that people from other cultures can have different values.", "Respecting other cultures is something that

students should learn as early as possible.”, “In the classroom, it is important that students of different origins recognize the similarities that exist between them.”, “When there are conflicts between students of different origins, they should be encouraged to resolve the argument by finding common ground.”. The values of the SCMCEG variable were transformed according to the WLE.

3.3.1.3 Criterion variables

Scientific literacy (PVSCIE)

Minimum: 0; Maximum: 1000; Scale of Measurement: Scale; $\alpha = 0.86$

Student’s Scientific literacy (PV1SCIE, PV2SCIE, PV3SCIE, PV4SCIE, PV5SCIE, PV6SCIE, PV7SCIE, PV8SCIE, PV9SCIE, PV10SCIE) was measured on 115 items in PISA2018. The items themselves are not available to the public for PISA2018. Students were given different subsets of the cognitive PISA test. Using the Rasch model (OECD, 2009b) student’s performances were estimated as if they took the whole test. In total 10 Plausible Values were computed for each individual student.

3.4 Instruments

For the assessment of education in Bosnia and Herzegovina in PISA2018 three instruments were used: Cognitive assessment instrument, Student questionnaire and School questionnaire. Teacher questionnaire was not administered, which could have been useful for this study as well. Responsibilities of organizing and administration of the tests were handled by APOSO (Agency for Preschool, Primary and Secondary Education).

Cognitive assessment lasted 120 minutes, out of which 60 minutes were devoted to Reading, and then 30 minutes each for Mathematics and Science. Student questionnaire was administered for an additional 35 minutes. School questionnaire was administered to schools’ principals. Since PISA measures trends in education, Cognitive questionnaires items are confidential and can’t be accessed from the public. Student and school questionnaires can be freely accessed and can be downloaded from the OECD website.

3.4.1 Quality monitoring, Validity and Reliability

“Specific standards underlie the PISA questionnaire development process as well as the implementation of the material into the final instruments. These standards aim at quality assurance as well as comparability of the data across countries/economies. Mechanisms for PISA 2018 included a review by the countries/economies, cognitive labs, linguistic translatability assessment, centralized transfer of trend material, negotiation of adaptations and linguistic verification. “ (OECD, 2019c) More details about the field trials, test design etc. can be accessed here (OECD, 2019b). Cronbach alpha coefficients are presented for all the variables in this study that needed reliability checks in the Section 3.3.

3.5 Procedures

Primary goal of the present study was to make a deeper analysis of PISA2018 results in the context of scientific literacy. Since such studies were not done earlier for Bosnia and Herzegovina, literature review was done to see which approaches and statistical methods were used to describe the relationship among variables PISA2018 measured and scientific literacy scores.

Some type of ANOVA or regression analysis seemed like the best approach. First, we decided on Structural Equation Modelling (SEM), since there were examples of it in the literature and understanding AMOS as a software was simple enough. However deeper analysis of the dataset revealed that for the nested structure of data Multilevel modeling (MLM) and precisely Hierarchical Linear Modelling (HLM) would be a better approach. Multilevel Structural Equation modelling was also used in some studies (Brühwiler & Blatchford, 2011); however, it proved too laborious. After deciding on HLM, we spent 2 months learning about the technique itself and then 2 months more familiarizing ourselves with the HLM7 software. HLM7 had a great advantage of an option to work with Plausible Values and making the whole process easier. One big downside was that it is a proprietary software and we were able to use it with the courtesy of prof Ulas Üstün through TeamViewer.

Next big obstacle came with the chosen model not converging. Chosen model was quite big, compared to the other in literature review (Üstün et al., 2020). The main reason for working with such a big number of predictor variables was no prior general models done for PISA in Bosnia

and Herzegovina and it also being in accordance with the Supply – Use theoretical model. The non – convergence issue was resolved with the “bottom – up” Analysis Strategy recommended by (J. J. Hox et al., 2017).

3.5.1 Analysis strategy

The strategy analysis chosen for this study is the so called “bottom – up” analysis strategy suggested by (J. J. Hox et al., 2017) in comparison to the “top – down” strategy suggested by (Raudenbush & Bryk, 2002). With models of this size, computational issues due to convergence errors are not rare. In addition, there was an interest in making a wide, rather than deep analysis of as much of data as possible. Even though the model does have a theoretical background it is still not rigid in the number of included variables, therefore it relies on statistics as well as the theory to find a good fit of a model. The analysis strategy consists of the following five steps:

Step 1: In this step, model is analyzed without predictor variables. Such model is named intercept – only model and it can be described by the following equation:

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \quad (2.1)$$

(γ_{00} is the regression intercept, or simply the mean of Y_{ij} in the sample, whereas u_{0j} and r_{ij} are just the usual residuals at the group and the individual level respectively).

Biggest usefulness of this model lies in the estimation of the intra – class correlation (ICC) given by: $ICC = \frac{\tau}{\sigma^2 + \tau}$. ICC is the percentage of between – group variance compared to the total variance. More information about ICC is given in the Section 4.3.2.

Step 2: In this step, all lower – level (i.e., student – level) predictor variables fixed (i.e., variance of the slopes is fixed to zero) are added and the model is analyzed. This model can be described by the following equation:

$$Y_{ij} = \gamma_{00} + \gamma_{p0} * X_{pij} + u_{0j} + r_{ij} \quad (2.2)$$

(X_{pij} are the p predictor variables at the student – level). Statistical significance of each predictor variable is assessed and between – school and within – school variances are noted.

Step 3: In this step, all higher – level (i.e., school – level) predictor variables are added and the model is analyzed. This model can be described by the following equation:

$$Y_{ij} = \gamma_{00} + \gamma_{p0} * X_{pij} + \gamma_{0q} * Z_{qj} + u_{0j} + r_{ij} \quad (2.3)$$

(Z_{qj} are the q predictor variables at the school – level). Statistical significance of each added predictor is assessed and between – school and within – school variances are noted.

The models in Step 2 and Step 3 are usually referred to as the *variance component* models, since they decompose the intercept variance into between – group and within – group variance for both levels in HLM.

Step 4: In this step, lower – level slope variances are assessed for significance. For this study, assessment was done visually (Appendix B). Variables which have significant slope variances between groups are let to vary in the model. Such model is named random – coefficient model and it can be described by the following equation:

$$Y_{ij} = \gamma_{00} + \gamma_{p0} * X_{pij} + \gamma_{0q} * Z_{qj} + u_{pj} * X_{pij} + u_{0j} + r_{ij} \quad (2.4)$$

(u_{pj} are the group – level residuals of the slopes of the individual – level explanatory variables X_{pij}). Statistical significance of each added predictor is assessed and between – school and within – school variances are noted.

Step 5: In this step, cross – level interactions among school – level variables and those student – level variables that had significant slope variation in the previous step are added. Such model is named *full model* and it can be described by the following equation:

$$Y_{ij} = \gamma_{00} + \gamma_{p0} * X_{pij} + \gamma_{0q} * Z_{qj} + \gamma_{0q} * X_{pij} * Z_{qj} + u_{pj} * X_{pij} + u_{0j} + r_{ij} \quad (2.5)$$

Statistical significance of each added predictor is assessed and between – school and within – school variances are noted.

The process of modeling at each step is done by assessing the statistical significance of each variable. There are many ways to assess the statistical significance, however in this thesis condition was set to $p < .05$. Coefficient of the variables that fulfilled that condition were kept in the analysis at each step, and others were excluded. Exclusion of the variables was done by backwards elimination. After running the model, the variable that had the biggest coefficient that was not statically significant would get excluded and model would rerun. The process continues until only statistically significant variables remain.

As a method of computation Full Maximum Likelihood (FML) was used. Student – level variables were group – mean centered, whereas school – level variables were grand – mean centered. Further discussion about possible choices can be found in (J. J. Hox et al., 2017; Raudenbush & Bryk, 2002).

3.5.2 Assumptions for the Hierarchical Linear Modeling (HLM)

Measuring the world in a quantitative way is quite hard. Only one of the reasons for that being that sample can never fully encompass all the intricacies of the real population. The other obstacle is that for any inferential statistics there is a set of assumptions that need to be fulfilled so that the subsequent analyses can be reliable. Multilevel modeling (MLM) and Hierarchical Linear Modeling (HLM) as a type of the previous, have the same assumptions as other general linear models like ANOVA and single-level regression with some additions considering additional predictor levels.

HLM is a powerful but complex statistical technique, for which there are still many unknowns considering failure of fulfillment of the assumptions. According to (J. J. Hox et al., 2017) and (Tabachnick & Fidell, 2014), assumptions that need to be fulfilled for the HLM analysis are as follows:

1. *Random Sampling*
 - i. Random sampling for level-1,
 - ii. Random sampling for level-2.
2. *Independence of Observation*

Note: Independence of observation, even though legitimate assumption of ordinary linear regression, cannot be fulfilled in HLM. Nested data, i.e., assumption that students nested within the same schools share some of the same characteristics is the main reason of choosing a type of multilevel modeling instead of single-level modeling. HLM as a model expects that independence of observation won't be fulfilled within schools and is robust to that assumption. (Tabachnick & Fidell, 2014)

3. *Sufficient sample size,*

Complex structure of the nested sample also warrants dealing with it in a more sensitive way than if it was one level sampling. If the focus is on the fixed effects, then the 50/20 rule is recommended, i.e., at least 50 groups with at least 20 participants per group. However if there is interest in random effects as well, then 100/10 rule is recommended, i.e. at least 100 groups with at least 10 participants. (J. J. Hox et al., 2017)

4. *Multicollinearity and Singularity,*

There is no significant ($r = .7$) bivariate correlation among predictor variables on both levels.

5. *Outliers,*

No significant outliers for both predictor and criterion variables. Outlier is defined as a value of standardized residual outside of the $[-3, +3]$ interval. If potential outlier is found, its influence can be evaluated by the Cook's Distance statistics. If the Cook's Distance is < 1 there is no substantial influence on the regression equation and model can be run without deletion of the outliers.

6. *Normality, Linearity, Homoscedascity of Residuals,*

- i. Criterion variables are multivariate normal,
- ii. Level-1 residuals are independent and multivariate normal,
- iii. Level-2 residuals are independent multivariate normal,

7. Independence of Residuals:

- i. Level-1 residuals are independent of the residuals at Level-2,
- ii. Predictors at level-1 are independent of the residuals at level-1,
- iii. Predictors at level-2 are independent of the residuals at level-1,

4. RESULTS

In this chapter, data analysis process is described as well as the results that were achieved. The chapter begins with the data cleaning and missing data analysis, continuing with descriptive statistics of level-1 and level-2 variables. In the final part of the chapter, summary of the results is presented starting with checking the statistical assumptions and in accordance to the posed research problems.

4.1. Data Cleaning and Missing Data Analysis

Preparation of the final versions of two datasets to be used in the model was done through data cleaning, missing data analysis, and eventually missing data replacement. The process was done in the same way for level-1 and level-2 files in the SPSS22 software, which are presented as follows:

Data were dealt with in a systematic and methodological way as to provide the reproducibility of the study. Two files, one for student questionnaire data (CY07_MSU_STU_QQQ) and another one for school questionnaire data (CY07_MSU_SCH_QQQ) were downloaded from the OECD website (<https://www.oecd.org/pisa/data/2018database/>) in the SPSS format and split according to countries (*CNTRYID*: 70 for BA), isolating data related to Bosnia and Herzegovina.

Student questionnaire dataset of PISA2018 for Bosnia and Herzegovina, from now on referred to as *level-1* data consisted of 6480 rows representing students and 1118 columns representing variables measured on that level. School questionnaire dataset of PISA2018 for Bosnia and Herzegovina, from now on referred to as *level-2* data consisted of 213 rows representing schools and 195 columns representing variables measures on that level. The procedure of data cleaning and missing data analysis proceeded as follows:

1. Descriptive analysis of all the variables showed that 642 variables did not have any data ($N = 0$) for level-1 dataset and there were no empty variables for the level-2 dataset. The empty variables were deleted.
2. In the following set of steps, based on the objective of the study of science literacy PISA scores, all unrelated variables were deleted, as so:

- a. All Plausible values except those related to science literacy (PV1SCIE, ..., PV10SCIE) were deleted,
 - b. All individual questions were deleted since this study deals only with indices,
 - c. All replicate weights (80) were deleted, since the software used for this study did not support its use,
 - d. All reading literacy domain specific indices were deleted,
3. Column-wise deletion was performed for the variables that had missing more than 30% of the data. That resulted in deleting following variables:
 - a. From level-1 dataset:
 - i. TMINS - Learning time (minutes per week) - in total (48%),
 - ii. DURECEC - Duration in early childhood education and care (63%).
 - b. From level-2 dataset:
 - i. SCHSIZE - School Size (Sum) (31%),
 - ii. STRATIO – Student-Teacher ratio (31%),
 - iii. RATCMP2 - Proportion of available computers that are connected to the Internet (44%).
4. Values of the variable SMINS had some values which did not make sense. All cases that had the value bigger than 450 were deleted (less than 5% of the sample). SMINS measures weekly minutes that students spend learning science (weekly classes multiplied by average duration of the class) as explained in Section 3.3.1.1. In the PISA2018 Bosnia and Herzegovina sample, the highest sampled grade was second grade of high-school. National curriculum of Bosnia and Herzegovina allocates 8 classes per week (as in the first grade of secondary school, 2 classes each: Physics, Biology, Chemistry and Geography) with an additional possible 2 facultative classes (maximum of 450 minutes is the maximum that students spend learning science per the SMINS definition).
5. IMMIG (Index Immigration Status) was recoded (3 = 2, 2 = 1, 1 = 0) into IMMIGRCD. ST004D01T (Student Gender Standardized) was recoded (2 = 1, 1 = 0) into GENDERRCD. The reason for recoding was to help the interpretability of the results.
6. At the final step Expectation Maximization (EM) of missing data replacement was performed in SPSS22 to keep as much of data as possible (IBM, 2019).

4.2 Descriptive Statistics

Descriptive statistics for level-1 (student level) as well as for level-2 (school level) variables are presented in Table 4.1 and 4.2, respectively. For continuous variables, measures of central tendency are given with Mean and Mode, measures of dispersion with minimum, maximum, and standard deviation (SD), and finally measures of distribution with skewness and kurtosis, with histograms. For nominal and one ordinal variable frequency as well as possible values of the variables are presented below. Correlation matrix for continuous variables for both levels are presented in Table 4.3.

Student Gender Recoded (GENDERECD) is a nominal dummy variable with the following values: 0 = "Female" (f = 3082), 0 = "Male" (f = 3235). Grade Repetition (REPEAT) is a nominal dummy variable with the following values: 0 = "Did not repeat a grade" (f = 6207), 1 = "Repeated a grade" (f = 110). Index Immigration Status Recoded (IMMIGRCD) is an ordinal dummy variable with the following values: 0 = "Native student" (f = 6145), 1 = "Second-generation student" (f = 83), 1 = "First-generation student" (f = 87).

Histograms of all continuous variables are presented in the Appendix A. GRADE and BLONG from level one, while CLSIZE, PROAT5AM, and PROAT6 from level two, have values of skewness and kurtosis outside of the (-2, +2) interval expected for univariate normal distribution. Histograms however indicate some normality. Bivariate correlations are given in Table 4.3 showing no significant correlation bigger than 0.61.

Table 4.1 Descriptive statistics for level-1 variables

LEVEL-1							
VARIABLES	Mean	Median	Min	Max	SD	Kurtosis	Skewness
GRADE	-0.15	0.00	-3.00	1.00	0.37	3.03	-2.01
AGE	15.79	15.75	15.33	16.33	0.28	-1.10	0.12
BSMJ	59.73	61.18	11.56	94.56	20.62	-1.07	-0.32
SMINS	147.32	120.00	0.00	450.00	97.60	0.94	1.01
ESCS	-0.56	-0.66	-4.79	3.42	0.82	0.11	0.28
EMOSUPS	-0.05	-0.18	-2.45	1.34	0.95	-0.88	-0.30
PERCOMP	0.24	0.24	-1.99	2.04	0.91	0.13	0.03

PERCOOP	0.02	-0.02	-2.14	1.68	0.97	-0.59	-0.09
ATTLNACT	0.00	0.03	-2.54	1.08	0.98	-0.71	-0.44
COMPETE	0.02	0.10	-2.35	2.01	1.01	0.21	0.08
WORKMAST	0.08	-0.10	-2.74	1.82	0.99	0.20	-0.05
GFOFAIL	-0.36	-0.38	-1.89	1.89	0.91	0.07	0.24
EUDMO	0.32	0.26	-2.15	1.74	0.91	0.18	-0.28
SWBP	0.35	0.50	-3.07	1.58	0.94	0.37	-0.94
RESILIENCE	0.23	-0.06	-3.17	2.37	1.02	0.65	0.31
MASTGOAL	0.20	0.22	-2.53	1.95	1.06	-0.35	-0.19
DISCRIM	0.23	0.32	-1.15	3.18	0.99	-0.05	0.33
BELONG	-0.05	-0.30	-3.26	2.78	0.94	2.42	1.03
BEINGBULLIED	-0.03	-0.78	-1.21	3.86	0.97	1.68	1.35
PV1SCIE	398.04	395.50	152.33	663.90	75.59	-0.22	0.16
PV2SCIE	398.47	395.53	146.99	657.02	75.61	-0.16	0.16
PV3SCIE	397.50	395.14	147.12	665.04	76.19	-0.16	0.13
PV4SCIE	398.76	396.06	161.19	682.09	76.92	-0.06	0.16
PV5SCIE	397.54	393.98	131.96	703.52	76.33	-0.08	0.21
PV6SCIE	398.94	395.37	140.78	681.24	76.76	-0.20	0.16
PV7SCIE	397.20	394.19	153.43	666.54	76.26	-0.23	0.12
PV8SCIE	398.64	395.31	137.06	688.10	76.28	-0.22	0.14
PV9SCIE	398.19	396.76	143.08	684.23	75.65	-0.12	0.13
PV10SCIE	398.13	395.22	161.47	676.09	76.88	-0.17	0.21

Note: Descriptive statistics given in this table are given for N = 6317 students. Missing entries were replaced with the Expectation Maximization (EM) technique as mention in the first part of this chapter, whereas descriptive statistic of dataset without missing data replacement is given in Appendix A.

Table 4.2 Descriptive statistics for level-2 variables

LEVEL-2							
VARIABLES	Mean	Median	Min	Max	SD	Kurtosis	Skewness
RATCMP2	0.75	1.00	0.00	1.00	0.35	-0.39	-1.07

TOTAT	44.58	39.00	11.00	136.50	20.75	1.99	1.21
PROATCE	0.79	1.00	0.00	1.00	0.37	0.33	-1.46
PROAT5AB	0.76	0.90	0.00	1.00	0.31	1.17	-1.55
PROAT5AM	0.07	0.04	0.00	1.00	0.10	36.11	4.95
PROAT6	0.01	0.00	0.00	0.07	0.01	6.85	2.62
CLSIZE	22.56	23.00	13.00	53.00	5.02	11.56	2.05
EDUSHORT	0.82	0.83	-1.42	2.96	1.00	-0.32	-0.02
STAFFSHORT	-0.72	-0.78	-1.47	2.09	0.77	0.14	0.79
STUBEHA	0.58	0.56	-4.03	3.58	1.42	0.52	-0.19
TEACHBEHA	0.19	0.23	-2.04	3.79	1.07	0.15	-0.11
SCMCEG	0.16	0.90	-4.59	1.29	1.13	1.92	-1.30

Note: Descriptive statistics are given for N = 213 schools. Missing entries were replaced with the Expectation Maximization (EM) technique, whereas descriptive statistic of dataset without missing data replacement is given in Appendix A.

Table 4.3 Bivariate correlations of the level-1, level-2 predictor variables and one (PV1SCIE) criterion variable.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
2	.42"																															
3	.02	.00																														
4	-.07"	-.06"	.24"																													
5	.06"	-.01	.31"	.16"																												
6	.02	.00	.19"	.07"	.11"																											
7	-.03"	.00	.05"	.03*	.04"	.31"																										
8	.00	.01	.06"	.03*	.04"	.33"	.37"																									
9	.02	.01	.11"	-.02	-.01	.21"	.13"	.18"																								
10	-.03*	-.01	.02	.01	.02	.14"	.23"	.11"	.03*																							
11	.00	.01	.17"	.05"	.07"	.35"	.27"	.25"	.18"	.38"																						
12	.03*	.01	-.02	-.02	-.03**	-.07"	.06"	-.01	.02	.09"	.02																					
13	.00	-.01	-.04"	-.01	.02	.28"	.22"	.23"	.14"	.23"	.35"	-.06"																				
14	-.01	-.02	.02	-.02	.02	.23*	.16"	.23*	.13"	.07"	.18"	-.12"	.37"																			
15	.01	.00	.13"	.05"	.10"	.37"	.30"	.25"	.13"	.31"	.46"	-.09"	.40"	.26"																		
16	-.01	.00	.14"	.04"	.05"	.35"	.28"	.30"	.36"	.17"	.39"	.04"	.30"	.27"	.37"																	
17	-.04"	-.03*	.18"	-.05"	-.04"	.15"	.08"	.10"	.14"	.07"	.11"	.11"	.01	-.06"	-.06"	-.09"																
18	.01	-.03*	.11"	.05"	.07"	.30"	.18"	.26"	.16"	.14"	.27"	.19"	.26"	.28"	.32"	.22"	.18"															
19	-.05"	-.02	-.14"	-.03"	-.05"	-.20"	-.01	.10"	.10"	.03*	.11"	.15"	-.06"	.19"	.10"	.10"	.29"	.26"														
20	.01	.00	.02	.01	.04"	-.03*	-.01	.00	-.02	.00	-.02	.03"	.02	-.01	-.02	-.02	.01	-.02	.03*													
21	.01	-.07"	.08"	.06"	.07"	.03*	.00	.00	.01	.00	.02	.00	.01	.00	.03*	.01	-.03*	.02	-.02	.04"												
22	-.07"	-.01	-.03"	-.04"	-.03*	.00	.01	.02	.00	.02	-.01	-.01	.02	.00	-.01	.00	.00	.00	.00	.00	.02	.03*										
23	.01	.01	.02	-.01	.01	.02	.00	.00	.01	.02	.02	.00	-.01	.03*	.01	.01	-.02	.02	-.02	.00	-.07"											
24	.07"	.06"	.01	.00	.02	.00	-.02	.01	.01	.01	.00	.00	-.02*	.02	.00	.01	-.02	.03*	-.03*	.00	-.14"	.00	.41"									
25	.07"	.02	.13"	.10"	.11"	.05"	.02	.04"	.04"	-.01	.04"	.00	.00	-.03*	.03*	.03*	-.02	.02	.05"	-.01	.10"	.13"	-.01	.00								
26	.17"	.06"	.03*	.05"	.07"	-.01	.02	.01	.01	.00	.02	.03*	-.01	-.06"	-.01	-.01	-.01	-.02	.03*	.03*	.01	-.03**	-.06"	.02	.20"							
27	.01	.03*	.13"	.05"	.05"	.06"	.03*	.04"	.02	-.02	.04"	-.02	-.01	-.01	.01	.03*	-.05"	.03*	-.02	.01	.07"	.04"	-.02	.08"		.17"						
28	-.17"	-.06"	.12"	.08"	.09"	.04"	.01	.01	.00	.00	.01	-.01	.00	.00	.01	.02	-.04**	.02	-.03*	.00	.17"	.18"	.01	-.01	.16"	-.01	-.02					
29	-.04"	-.03*	-.03**	-.02	-.07**	.00	-.01	.00	.03*	.01	-.02	.02	.02	.00	.01	.03*	.03**	.01	.00	.05**	-.08**	.04**	-.11"	-.10"	.01	-.13"	-.09"	.02				
30	-.03*	-.02	-.06"	-.05**	.01	.01	.02	.02	.03**	.03*	.00	.01	.03**	.01	.03*	.02	.01	.00	.01	-.02	-.04**	.00	-.05"	-.10"	-.02	-.05"	.03*	.26"				
31	.24"	.09"	-.02	-.03*	-.01	-.01	.00	.01	.02	.02*	.00	.05**	.03**	-.04**	.00	.01	.04**	-.02	.03*	.03**	.00	.00	-.03*	-.03*	-.028*	.19**	-.07**	.23**	.06**			
32	.15"	.03*	.07**	.03*	.05**	.04**	.03*	.04**	.03*	.02	.02*	.01	.03*	.00	.02	.03*	.01	.02	-.03**	.06**	.04**	-.10"	.16"	.05**	.02*	-.03*	-.02	.28**	.27**	.60**		
33	-.04"	-.03**	.06**	.01	.03*	.02	-.02	-.01	-.01	-.01	-.01	.00	-.02	-.01	.00	-.01	-.03**	.02	-.03*	.02	.05**	.03*	.09**	.05**	.03*	.11**	.13**	-.16**	-.20**	-.10**	-.19**	
34	.08"	.02	.44**	.30**	.27**	.22**	.06**	.08**	.01	.06**	.19**	-.06**	-.03*	-.02	.14**	.08**	.19**	.14**	-.15**	-.06**	.07**	.01	.04**	.04**	.12**	.03*	.09**	.14**	-.07**	-.05**	-.03**	.08**

The number indicates: 1 (GRADE), 2 (AGE), 3 (BSMJ), 4 (SMINS), 5 (ESCS), 6 (EMOSUPS), 7 (PERCOMP), 8 (PERCOOP), 9 (ATLNACT), 10 (COMPETE), 11 (WORKMAST), 12 (GFOFAIL), 13 (EUDMO), 14 (SWBP), 15 (RESILIENCE), 16 (MASTGOAL), 17 (DISCRIM), 18 (BELONG), 19 (BEINGBULLIED), 20 (IMMIGRD), 21 (RATCMP2), 22 (TOTAT), 23 (PROATCE), 24 (PROAT5AB), 25 (PROAT5AM), 26 (PROAT6), 27 (CLSIZE), 28 (CREACTIV), 29 (EDUSHORT), 30 (STAFFSHORT), 31 (STUBEHA), 32 (TEACHBEHA), 33 (SCMCEG), 34 (PV1SCIE).

Note: * indicates p<.05; ** indicates p<.01.

4.3 Inferential Statistics

This section describes the process of checking the underlying assumptions for HLM and then presents the results of analyses that leads to the answer of the five research questions posed in this thesis. Each research question is answered with a certain model, for which the coefficients are presented. Starting with the Null model to determine base values of variances, continuing onto the *variance component models* that decompose level-1 and level-2 variances separately. Finally, by analyzing Random coefficient model and the Final model, taking both levels into consideration as well as appropriate randomness in the slopes, final model is presented.

4.3.1 Verification of the HLM assumptions

1. (*Random sampling*) OECD reports stratified random sampling as the procedure of gathering the data for PISA. Assumption is fulfilled.
2. (*Independence of Observation*) HLM is robust and specifically created to take into consideration lack of independence of observation. Assumption is fulfilled.
3. (*Sufficient sample size*) A sample size of 6317 level-1 and 213 level-2 is sufficient for the sample size assumption to be fulfilled, as indicated, and calculated in Section 3.5.3.
4. (*Multicollinearity and Singularity*) As it can be seen in Table 4.3 there is no significant bivariate correlation larger than 0.61 across and between level-1 and level-2 predictors as well as the criterion variable (PV1SCIE). Assumption is satisfied.
5. (*Outliers*) There were some outliers, like the ones for SMINS, which were removed from the study. In addition, Cook's distance was less than 1, meaning that remaining outliers do not significantly affect the analysis, which means we do not need to exclude them from the analysis. Assumption is satisfied.
6. (*Normality, Linearity, Homoscedasticity of Residuals*)
 - a. Univariate normality was checked in the descriptive statistics. Criterion variable and most of the level-1 and level-2 predictors were normally distributed. Multivariate normality was not checked. It is important to say that HLM is robust to non-normality, especially with such a big sample, Maximum Likelihood estimates will not be biased (J. Hox, 1998).

- b. The following assumption as checked according to the hierarchical multiple linear regression assumption. Scatterplot made from both level predictors nested within each other, and one of the plausible values for criterion variable is portrayed in Figure 4.1. Approximately rectangular shape of the points indicates that standard deviation of errors of prediction can be considered equal for predicted Science literacy plausible value (PV1SCIE), fulfilling homoscedasticity assumption. Assumption of linearity is fulfilled as well as can be concluded from linear relationship between predicted value and residuals. In addition, residual scores pilling around zero and having symmetrical distribution around the zero point fulfils the normality assumption (Tabachnick & Fidell, 2014). The normal Q-Q plot also further confirm normality as seen as seen in Figure 4.2.

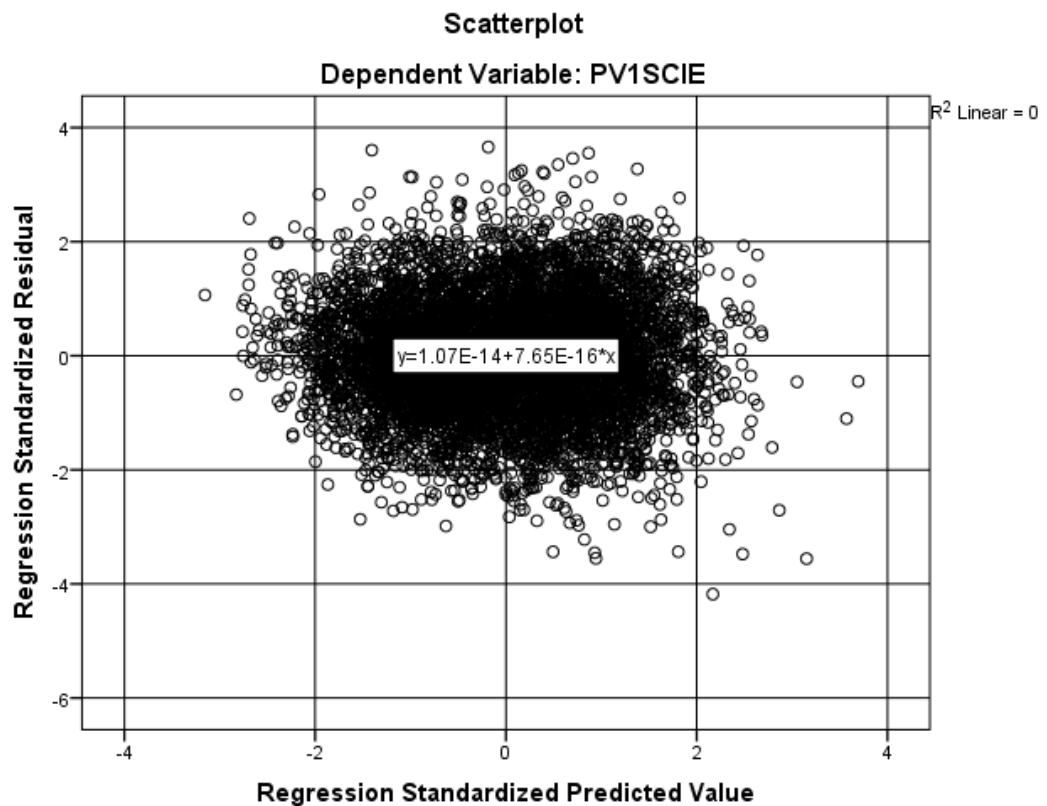


Figure 4.1 Scatterplot of the regression residuals to the predicted value

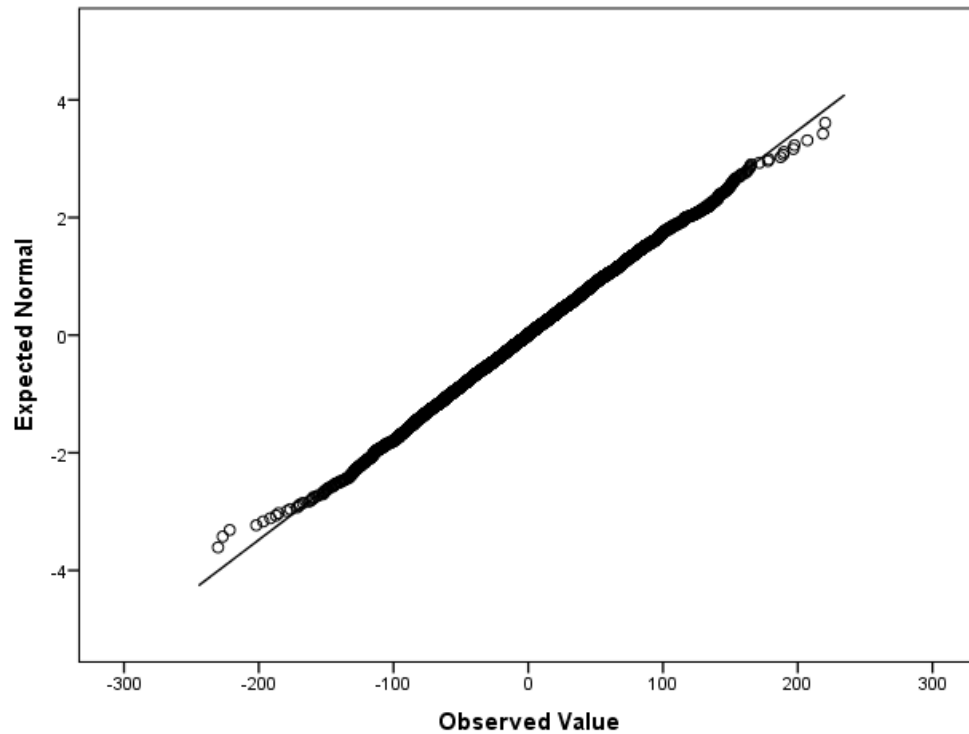


Figure 4.2 Normal Q-Q plot of the residuals

7. (*Independence of residuals*) Durbin-Watson statistics of the value of 1.782 which is between 1.5 and 2.5 satisfies this assumption.

Even though some assumptions, like normality, were not ideally satisfied, due to the robustness of HLM, size of the sample and scope of this study, assumptions can be considered as fulfilled to a satisfactory degree.

4.3.2 Results for Research Question I (Null model)

Research Question I: *What are between-school and within-school variances in PISA2018 scientific literacy scores of 15-year-old students in Bosnia and Herzegovina?*

The answer to this research questions was given in the null model, presented in Table 4.5. Null, empty, unconditional, or intercept-only model are all different names for the same model. In this model, we have a chance to determine base within-group (σ^2) and between-group (τ) variances. From null model, intra-class correlation (ICC) can be calculated, using the equation: $ICC = \frac{\tau}{\sigma^2 + \tau}$. ICC is the percentage of between-group variance compared to the total variance. The value of ICC is also used to justify the use of HLM compared to the ordinary linear regression (Tabachnick & Fidell, 2014).

$$\text{Level-1 Model } PVSCIE_{ij} = \beta_{0j} + r_{ij}$$

$$\text{Level-2 Model } \beta_{0j} = \gamma_{00} + u_{0j}$$

$$\text{Mixed Model } PVSCIE_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

Where $PVSCIE_{ij}$ is the science literacy score (all plausible values taken into consideration) of the student i in the school j . β_{0j} is the intercept of γ_{00} which is science literacy grand mean of all students, u_{0j} is random error associated with school j and r_{ij} is random error associated with student i in school j .

Table 4.5 *Results of the Null Model*

	Coefficient	SE	t-ratio	Approx. d.f.	p-value
γ_{00} (Science literacy grand mean)	395.70	3.08	128.34	212	<.001

Table 4.6 *Null model for variance components*

	Variance	d.f.	χ^2	p-value
Between-school (τ)	1416.04	212	2577.69	<.001
Within-school (σ^2)	4238.04			
Total	5654.08			
Intra-class correlation (ICC)		0.25		

Summary of the null-model

Science literacy grand mean score of 395.70 (212, N = 213), $p < .001$ presents a benchmark value of science literacy coefficient pooled across schools, to which we will compare our upcoming regression coefficients. It could be useful to mention that student mean score is calculated and reported (Džumhur, 2019) as 398.5, however that score does not take nested nature of data into consideration. Between-school variation of $\tau = 1416.04$, within-school variation of $\sigma^2 = 4238.04$ and the value of ICC show that about 25% variation in the science literacy scores of students can be explained with the variation between the schools. Chi-square test results χ^2 (212, N = 213) = 2577.69, $p < .001$ show statistically significant between-school variation in science literacy score. Both ICC and chi-square results go in favor of HLM compared to ordinary linear regression.

4.3.3 Results of Research Question II (student-level model)

Research Question II: *Which student-level variables predict scientific literacy scores of 15-year-old students as measured in PISA2018 in Bosnia and Herzegovina?*

The answer to this research question was acquired in the second step of the previously described Analysis strategy, in which all student-level (level-1) predictors are being added to the fixed effects model. According to our theoretical model, initially 22 variables were added, however only 14 variables turned out to be statistically significant, as presented in Table 4.6.

Level-1 Model

$$PVISCIE_{ij} = \beta_{0j} + \beta_{1j}*(GRADE_{ij}) + \beta_{2j}*(REPEAT_{ij}) + \beta_{3j}*(BSMJ_{ij}) + \beta_{4j}*(SMINS_{ij}) + \beta_{5j}*(ESCS_{ij}) + \beta_{6j}*(EMOSUPS_{ij}) + \beta_{7j}*(ATTLNACT_{ij}) + \beta_{8j}*(WORKMAST_{ij}) + \beta_{9j}*(GFOFAIL_{ij}) + \beta_{10j}*(EUDMO_{ij}) + \beta_{11j}*(SWBP_{ij}) + \beta_{12j}*(DISCRIM_{ij}) + \beta_{13j}*(IMMIGRCD_{ij}) + \beta_{14j}*(GENDERRC_{ij}) + r_{ij}$$

Level-2 Model

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + u_{0j} \\ \beta_{1j} &= \gamma_{10} \\ \beta_{2j} &= \gamma_{20} \\ \beta_{3j} &= \gamma_{30} \\ \beta_{4j} &= \gamma_{40}\end{aligned}$$

$$\begin{aligned}\beta_{5j} &= \gamma_{50} \\ \beta_{6j} &= \gamma_{60} \\ \beta_{7j} &= \gamma_{70} \\ \beta_{8j} &= \gamma_{80} \\ \beta_{9j} &= \gamma_{90} \\ \beta_{10j} &= \gamma_{100} \\ \beta_{11j} &= \gamma_{110} \\ \beta_{12j} &= \gamma_{120} \\ \beta_{13j} &= \gamma_{130} \\ \beta_{14j} &= \gamma_{140}\end{aligned}$$

Mixed Model

$$PVISCIE_{ij} = \gamma_{00} + \gamma_{10} * GRADE_{ij} + \gamma_{20} * REPEAT_{ij} + \gamma_{30} * BSMJ_{ij} + \gamma_{40} * SMINS_{ij} + \gamma_{50} * ESCS_{ij} + \gamma_{60} * EMOSUPS_{ij} + \gamma_{70} * ATTLNACT_{ij} + \gamma_{80} * WORKMAST_{ij} + \gamma_{90} * GFOFAIL_{ij} + \gamma_{100} * EUDMO_{ij} + \gamma_{110} * SWBP_{ij} + \gamma_{120} * DISCRIM_{ij} + \gamma_{130} * IMMIGRCD_{ij} + \gamma_{140} * GENDERRC_{ij} + u_{0j} + r_{ij}$$

Table 4.6 Results of the Fixed effects model for level-1 variables

Fixed Effect		Coefficient	SE	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0						
INTRCPT2, γ_{00}		395.59	3.08	128.41	212	<.001
GRADE, β_1 , γ_{10}		35.10	14.71	2.39	28	.024
REPEAT, β_2 , γ_{20}		-37.19	10.98	-3.39	18	.003
BSMJ, β_3 , γ_{30}		0.99	0.06	16.37	149	<.001
SMINS, β_4 , γ_{40}		0.12	0.01	10.44	65	<.001
ESCS, β_5 , γ_{50}		4.00	1.36	2.94	158	.004
EMOSUPS, β_6 , γ_{60}		8.55	1.68	5.08	18	<.001
ATTLNACT, β_7 , γ_{70}		-2.68	1.05	-2.55	86	.012
WORKMAST, β_8 , γ_{80}		8.52	1.03	8.26	128	<.001
GFOFAIL, β_9 , γ_{90}		-2.93	1.14	-2.57	64	.013
EUDMO, β_{10} , γ_{100}		-6.04	1.25	-4.85	77	<.001
SWBP, β_{11} , γ_{110}		-4.54	1.33	-3.42	57	.001
DISCRIM, β_{12} , γ_{120}		-5.99	1.27	-4.72	29	<.001
IMMIGRCD, β_{13} , γ_{130}		-16.33	4.27	-3.82	36	<.001
GENDERRC, β_{14} , γ_{140}		22.22	2.19	-10.13	123	<.001

Within-school variance (σ^2) = 3318.53

Between-school variance (τ) = 1450.11

Note: Level-1 (student level) variables were Group mean centered before the analysis.

Summary of the level-1 model:

Results from level-1 model presented in Table 4.6 indicate significant influence of: *Grade compared to modal grade in country* (GRADE), ($\gamma_{10} = 35.10$, SE = 14.71); *Grade Repetition* (REPEAT), ($\gamma_{20} = -37.19$, SE = 10.98); *Student's expected occupational status* (BSMJ), ($\gamma_{30} = 0.99$, SE = 0.06); *Learning time (minutes per week) - <science>* (SMINS), ($\gamma_{40} = 0.12$, SE = 0.01); *Index of economic, social and cultural status* (ESCS), ($\gamma_{50} = 4.00$, SE = 1.36); *Parents' emotional support perceived by student* (EMOSUPS), ($\gamma_{60} = 8.55$, SE = 1.68); *Attitude towards school: learning activities* (ATTLNACT), ($\gamma_{70} = -2.68$, SE = 1.05); *Work mastery* (WORKMAST), ($\gamma_{80} = 8.52$, SE = 1.03); *Eudaemonia: meaning in life* (EUDMO), ($\gamma_{100} = -6.04$, SE = 1.25); *Subjective well-being* (SWBP), ($\gamma_{110} = -4.54$, SE = 1.33); *Discriminating school climate* (DISCRIM), ($\gamma_{120} = -5.99$, SE = 1.27); *Index Immigration status Recoded* (IMMIG), ($\gamma_{130} = -16.33$, SE = 4.27); *Students' Gender Recoded* (GENDERRCD), ($\gamma_{140} = 22.22$, SE = 2.19) and *Science Literacy scores*. These 14 factors are reevaluated again in Section 4.3.5 Intercepts and Slopes as Outcomes Model.

Original within-school variance was 4238.03 and it dropped to 3318.53 after inclusion of the 14 level-1 predictors. It can also be seen that between-school variability slightly increased (from 1416.04 to 1450.11).

$$\text{Effect size: } R^2 = \frac{\sigma^2_{\text{Null model}} - \sigma^2_{\text{Fixed level-1 model}}}{\sigma^2_{\text{Null model}}} = .22$$

R-square value of .22 indicates medium effect size (Cohen, 2013). In other words, 22% of total within-school variance is attributable to the previously presented 14 level-1 variables.

4.3.4 Results of Research Question III (school-level model)

Research Questions III: *Which school-level variables predict scientific literacy scores of 15-year-old students as measured in PISA2018 in Bosnia and Herzegovina?*

The answer to this research question was acquired in the third step of the previously described Analysis strategy, in which all school (level-2) level predictors are being added to the fixed effects model of level-1 variables. The same result, with negligible differences in coefficient

values and variances can be achieved by running the Means as Outcomes model, which includes only variables from the level-2 (Raudenbush & Bryk, 2002).

According to our theoretical model, initially 13 variables were added to the level-1 fixed effects model, however only 5 variables turned out to be statistically significant, as presented in Table 4.7. without the fixed effects for level-1.

Level-1 Model

$$PVSCIE_{ij} = \beta_{0j} + \beta_{1j}*(GRADE_{ij}) + \beta_{2j}*(REPEAT_{ij}) + \beta_{3j}*(BSMJ_{ij}) + \beta_{4j}*(SMINS_{ij}) + \beta_{5j}*(ESCS_{ij}) + \beta_{6j}*(EMOSUPS_{ij}) + \beta_{7j}*(ATTLNACT_{ij}) + \beta_{8j}*(WORKMAST_{ij}) + \beta_{9j}*(GFOFAIL_{ij}) + \beta_{10j}*(EUDMO_{ij}) + \beta_{11j}*(SWBP_{ij}) + \beta_{12j}*(DISCRIM_{ij}) + \beta_{13j}*(IMMIGRCD_{ij}) + \beta_{14j}*(GENDERRC_{ij}) + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(PROAT5AM_j) + \gamma_{02}*(CLSIZE_j) + \gamma_{03}*(CREACTIV_j) + \gamma_{04}*(EDUSHORT_j) + \gamma_{05}*(TEACHBEH_j) + u_{0j}$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30}$$

$$\beta_{4j} = \gamma_{40}$$

$$\beta_{5j} = \gamma_{50}$$

$$\beta_{6j} = \gamma_{60}$$

$$\beta_{7j} = \gamma_{70}$$

$$\beta_{8j} = \gamma_{80}$$

$$\beta_{9j} = \gamma_{90}$$

$$\beta_{10j} = \gamma_{100}$$

$$\beta_{11j} = \gamma_{110}$$

$$\beta_{12j} = \gamma_{120}$$

$$\beta_{13j} = \gamma_{130}$$

$$\beta_{14j} = \gamma_{140}$$

Mixed Model

$$PVSCIE_{ij} = \gamma_{00} + \gamma_{01}*PROAT5AM_j + \gamma_{02}*CLSIZE_j + \gamma_{03}*CREACTIV_j + \gamma_{04}*EDUSHORT_j + \gamma_{05}*TEACHBEH_j + \gamma_{10}*GRADE_{ij} + \gamma_{20}*REPEAT_{ij} + \gamma_{30}*BSMJ_{ij} + \gamma_{40}*SMINS_{ij} + \gamma_{50}*ESCS_{ij} + \gamma_{60}*EMOSUPS_{ij} + \gamma_{70}*ATTLNACT_{ij} + \gamma_{80}*WORKMAST_{ij} + \gamma_{90}*GFOFAIL_{ij} + \gamma_{100}*EUDMO_{ij} + \gamma_{110}*SWBP_{ij} + \gamma_{120}*DISCRIM_{ij} + \gamma_{130}*IMMIGRCD_{ij} + \gamma_{140}*GENDERRC_{ij} + u_{0j} + r_{ij}$$

Table 4.7: Results of the school level (level-2) model

Fixed Effect		Coefficient	SE	t-ratio	Approx. d.f.	p-value
INTRCPT1, β_0						
INTRCPT2,	γ_{00}	395.51	2.82	140.08	207	<.001
PROAT5AM,	γ_{01}	78.40	19.16	4.09	207	<.001

CLSIZE,	γ_{02}	1.21	0.55	2.21	207	.028
CREACTIV,	γ_{03}	9.14	3.10	2.95	207	.004
EDUSHORT,	γ_{04}	-6.14	2.89	-2.12	207	.035
TEACHBEH,	γ_{05}	7.91	2.60	3.04	207	.003

Within-school variance (σ^2) = 3317.76

Between-school variance (τ) = 1191.27

Note: Level-2 (school level) variables were Grand mean centered before the analysis.

Summary of the level-2 model:

Results from level-2 model presented in Table 4.7 indicate significant association among *Index proportion of all teachers ISCED LEVEL 5A Master* (PROAT5AM), ($\gamma_{01} = 78.40$, SE = 19.16); *Class Size* (CLSIZE), ($\gamma_{02} = 1.21$, SE = 0.55); *Creative extra-curricular activities* (CREACTIV), ($\gamma_{03} = 9.14$, SE = 3.10); *Shortage of educational material* (EDUSHORT), ($\gamma_{04} = -6.14$, SE = 2.89); *Teacher behavior hindering learning* (TEACHBEH), ($\gamma_{05} = 7.91$, SE = 2.60) and Science literacy scores. These four factors are reevaluated again in Section 4.3.5 Intercepts and Slopes as Outcomes Model.

From the final estimation of variance components presented at the end of Table 4.7 whereas within school variance did not really change, (from 3318.53 to 3317.76) between-school variance did change significantly. Original between school variance was 1416.04, whereas the residual variance between school, after the inclusion of school level variables dropped to 1191.27.

$$\text{Effect size: } R^2 = \frac{\tau_{\text{Null model}} - \tau_{\text{Fixed level-2 model}}}{\tau_{\text{Null model}}} = 0.16$$

R-square value of .16 indicates medium effect size (Cohen, 2013). In other words, 16% of total between-school variance is attributable to the previously presented 5 level-2 variables.

4.3.5 Random coefficient model

Research Questions IV: *Which combination of school-level and student-level variables predicts scientific literacy scores of 15-year-old students measured in PISA2018 in Bosnia and Herzegovina?*

The answer to this research question was acquired though the fourth step of the previously described Analysis strategy. In addition to the already determined level-2 variables, slopes of level-1 continuous variables are allowed to vary, making this model the Random Coefficients model, presented in the Table 4.8.

Level-1 Model

$$PVSCIE_{ij} = \beta_{0j} + \beta_{1j}*(REPEAT_{ij}) + \beta_{2j}*(BSMJ_{ij}) + \beta_{3j}*(SMINS_{ij}) + \beta_{4j}*(ESCS_{ij}) + \beta_{5j}*(EMOSUPS_{ij}) + \beta_{6j}*(ATTLNACT_{ij}) + \beta_{7j}*(WORKMAST_{ij}) + \beta_{8j}*(GFOFAIL_{ij}) + \beta_{9j}*(EUDMO_{ij}) + \beta_{10j}*(SWBP_{ij}) + \beta_{11j}*(DISCRIM_{ij}) + \beta_{12j}*(IMMIGRCD_{ij}) + \beta_{13j}*(GENDERRC_{ij}) + r_{ij}$$

Level-2 Model

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01}*(PROAT5AM_j) + \gamma_{02}*(CLSIZE_j) + \gamma_{03}*(CREACTIV_j) + \gamma_{04}*(EDUSHORT_j) \\ &\quad + \gamma_{05}*(TEACHBEH_j) + u_{0j} \\ \beta_{1j} &= \gamma_{10} \\ \beta_{2j} &= \gamma_{20} + u_{2j} \\ \beta_{3j} &= \gamma_{30} + u_{3j} \\ \beta_{4j} &= \gamma_{40} + u_{4j} \\ \beta_{5j} &= \gamma_{50} + u_{5j} \\ \beta_{6j} &= \gamma_{60} + u_{6j} \\ \beta_{7j} &= \gamma_{70} + u_{7j} \\ \beta_{8j} &= \gamma_{80} \\ \beta_{9j} &= \gamma_{90} + u_{9j} \\ \beta_{10j} &= \gamma_{100} + u_{10j} \\ \beta_{11j} &= \gamma_{110} + u_{11j} \\ \beta_{12j} &= \gamma_{120} + u_{12j} \\ \beta_{13j} &= \gamma_{130}\end{aligned}$$

Mixed Model

$$\begin{aligned}PVSCIE_{ij} &= \gamma_{00} + \gamma_{01}*PROAT5AM_j + \gamma_{02}*CLSIZE_j + \gamma_{03}*CREACTIV_j + \gamma_{04}*EDUSHORT_j + \\ &\quad \gamma_{05}*TEACHBEH_j + \gamma_{10}*REPEAT_{ij} + \gamma_{20}*BSMJ_{ij} + \gamma_{30}*SMINS_{ij} + \gamma_{40}*ESCS_{ij} + \gamma_{50}*EMOSUPS_{ij} + \\ &\quad \gamma_{60}*ATTLNACT_{ij} + \gamma_{70}*WORKMAST_{ij} + \gamma_{80}*GFOFAIL_{ij} + \gamma_{90}*EUDMO_{ij} + \gamma_{100}*SWBP_{ij} + \\ &\quad \gamma_{110}*DISCRIM_{ij} + \gamma_{120}*IMMIGRCD_{ij} + \gamma_{130}*GENDERRC_{ij} + u_{0j} + u_{2j}*BSMJ_{ij} + u_{3j}*SMINS_{ij} + \\ &\quad u_{4j}*ESCS_{ij} + u_{5j}*EMOSUPS_{ij} + u_{6j}*ATTLNACT_{ij} + u_{7j}*WORKMAST_{ij} + u_{9j}*EUDMO_{ij} + \\ &\quad u_{10j}*SWBP_{ij} + u_{11j}*DISCRIM_{ij} + u_{12j}*IMMIGRCD_{ij} + r_{ij}\end{aligned}$$

Table 4.8: Results of the random effects model

Fixed Effect	Coefficient	SE	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					

INTRCPT2,	γ_{00}	395.48	2.82	139.85	207	<.001
PROAT5AM,	γ_{01}	84.97	20.98	4.05	112	<.001
CLSIZE,	γ_{02}	1.20	0.52	2.27	207	.024
CREACTIV,	γ_{03}	8.48	3.07	2.75	207	.006
EDUSHORT,	γ_{04}	-5.86	2.79	-2.09	207	.037
TEACHBEH,	γ_{05}	6.87	2.41	2.84	207	.005
For REPEAT slope, β_1						
INTRCPT2,	γ_{10}	-37.33	11.05	-3.38	17	.004
For BSMJ slope, β_2						
INTRCPT2,	γ_{20}	0.99	0.06	16.32	128	<.001
For SMINS slope, β_3						
INTRCPT2,	γ_{30}	0.12	0.01	10.19	56	<.001
For ESCS slope, β_4						
INTRCPT2,	γ_{40}	3.89	1.41	2.75	139	.007
For EMOSUPS slope, β_5						
INTRCPT2,	γ_{50}	8.60	1.68	5.11	19	<.001
For ATTNLNACT slope, β_6						
INTRCPT2,	γ_{60}	-2.57	1.04	-2.48	92	.015
For WORKMAST slope, β_7						
INTRCPT2,	γ_{70}	8.48	1.03	8.18	138	<.001
For GFOFAIL slope, β_8						
INTRCPT2,	γ_{80}	-3.12	1.17	-2.67	53	.010
For EUDMO slope, β_9						
INTRCPT2,	γ_{90}	-5.94	1.28	-4.63	65	<.001
For SWBP slope, β_{10}						
INTRCPT2,	γ_{100}	-4.44	1.30	-3.41	55	.001
For DISCRIM slope, β_{11}						
INTRCPT2,	γ_{110}	-5.98	1.26	-4.74	29	<.001
For IMMIGRCD slope, β_{12}						
INTRCPT2,	γ_{120}	-15.58	4.30	-3.62	38	<.001
For GENDERRC slope, β_{13}						
INTRCPT2,	γ_{130}	22.32	2.12	10.52	159	<.001
Within school variance (σ^2) = 3140.45						
Between school variance (τ) = 1204.20						

Summary of the Random Coefficient model:

In this model, GRADE ceased to be statistically significant, and was therefore excluded from the model. There were some changes in the values of level-1 and level-2 coefficients, however overall model did not change significantly from its fixed counterpart.

In this summary, more detail is put into description of the meaning of each coefficient, as to paint a more precise picture of these results. All the following explanation of points are meant in comparison to the grand mean (395.48).

Level-2 variables:

- *Index proportion of all teachers ISCED LEVEL 5A Master (PROAT5AM)*, ($\gamma_{01} = 84.97$, SE = 20.98) indicates how in school a unit increase in the value of PROAT5AM, students score on average 84.97 points more.
- *Class Size (CLSIZE)*, ($\gamma_{02} = 1.20$, SE = 0.52) indicates how a one unit increase in the average schools CLSIZE value, results in its students scoring on average 1.20 points more in science literacy.
- *Creative extra-curricular activities (CREACTIV)*, ($\gamma_{03} = 8.48$, SE = 3.07) indicates how a one unit increase in schools CREATIV value, results in its students score on average 8.48 points more in science literacy.
- *Shortage of educational material (EDUSHORT)*, ($\gamma_{04} = -5.86$, SE = 2.79) indicates how a one unit increase in a schools EDUSHORT value, results in its students having on average 5.86 points less in science literacy.
- *Teacher behavior hindering learning (TEACHBEH)*, ($\gamma_{05} = 6.87$, SE = 2.41) indicates how a one unit increase in a schools TEACHBEH value, results in its students having on average 6.87 points more in science literacy.

Level-1 variables:

- *Student's expected occupational status (BSMJ)*, ($\gamma_{20} = 0.99$, SE = 0.06) indicates how a one unit increase in students BSMJ value, results in that students score being on average 0.99 point bigger in science literacy.
- *Learning time (minutes per week) - <science> (SMINS)*, ($\gamma_{30} = 0.12$, SE = 0.01) indicates how a one-minute increase in students SMINS value, results in that students score being on average 0.12 point bigger in science literacy. The range of SMINS should also be considered (Min = 0, Max = 450) as to interpret the effect of this result. For example, a

student spending 459 minutes per week studying science, will have 49 points increase in their science literacy.

- *Index of economic, social, and cultural status (ESCS)*, ($\gamma_{40} = 3.89$, SE = 1.41) indicates how a one unit increase in students ESCS value, results in that students score being on average 3.89 point bigger in science literacy.
- *Parents' emotional support perceived by student (EMOSUPS)*, ($\gamma_{50} = 8.60$, SE = 1.68) indicates how a one unit increase in students EMOSUPS value, results in that students score being on average 8.60 point bigger in science literacy.
- *Attitude towards school: learning activities (ATTLNACT)*, ($\gamma_{60} = -2.57$, SE = 1.04) indicates how a one unit increase in students ATTLNACT value, results in that students score being on average 2.57 points smaller in science literacy.
- *Work mastery (WORKMAST)*, ($\gamma_{70} = 8.48$, SE = 1.03) indicates how a one unit increase in students WORKMAST value, results in that students score being on average 8.48 points bigger in science literacy.
- *General Fear of Failure (GFOFAIL)*, ($\gamma_{80} = -3.12$, SE = 1.17) indicates how a one unit increase in students GFOFAIL value, results in that students score being on average 3.12 points smaller in science literacy.
- *Eudaemonia: meaning in life (EUDMO)*, ($\gamma_{90} = -5.94$, SE = 1.28) indicates how a one unit increase in students EUDMO value, results in that students score being on average 5.94 points smaller in science literacy.
- *Subjective well-being (SWBP)*, ($\gamma_{100} = -4.44$, SE = 1.30) indicates how a one unit increase in students SWBP value, results in that students score being on average 4.44 points smaller in science literacy.
- *Discriminating school climate (DISCRIM)*, ($\gamma_{110} = -5.98$, SE = 1.26) indicates how a one unit increase in students DISCRIM value, results in that students score being on average 5.98 points smaller in science literacy.
- *Student's Gender Recoded (GENDERRCD)*, ($\gamma_{130} = 22.32$, SE = 2.12) indicates how males (GENDERRCD = 1) have on average 22.32 more points in science literacy compared to females (GENDERRCD = 0).
- *Index Immigration status (IMMIGRCD)*, ($\gamma_{120} = -15.58$, SE = 4.30) indicates how the students belonging to "Second-Generation" immigrants (IMMIGRCD = 1) have on average

15.58 less points whereas students belonging to “First-Generation” immigrants (IMMIGRCD = 2) have 32.36 (2*16.18) less points compared to the “Native” students (IMMIGRCD = 0).

As far as the between-school and within-school variances are concerned, interpreting them in this model is more complex, since slopes of the level-1 predictors are allowed to vary. It could be useful to mention however that allowing for the slopes to be random, had further decreased within-school variability ($\sigma^2 = 3140.45$) compared to the fixed model ($\sigma^2 = 3317.76$), with between school variability being slightly increased ($\tau = 1204.20$) from the previous step ($\tau = 1191.27$).

4.3.6 Full model (Interaction model)

Research Questions IV: *Are there any significant interactions between student-level and school-level variables that predict scientific literacy scores of 15-year-old students as measured in PISA2018 in Bosnia and Herzegovina have significant interactions?*

The answer to this research question was acquired through the fifth step of the previously described Analysis strategy. In the fifth step, interaction factors among five level-2 variables and each out of 13 level-1 variables with significant slope variation, were tested, yielding two statistically significant interactions, one between CREATIV and ESCS and other one between IMMIGRCD and PROAT5AM. Excluding all other insignificant interactions, model has been rerun resulting in the final model, presented in the Table 4.9.

Level-1 Model

$$PVSCIE_{ij} = \beta_{0j} + \beta_{1j}*(REPEAT_{ij}) + \beta_{2j}*(BSMJ_{ij}) + \beta_{3j}*(SMINS_{ij}) + \beta_{4j}*(ESCS_{ij}) + \beta_{5j}*(EMOSUPS_{ij}) + \beta_{6j}*(ATTLNACT_{ij}) + \beta_{7j}*(WORKMAST_{ij}) + \beta_{8j}*(GFOFAIL_{ij}) + \beta_{9j}*(EUDMO_{ij}) + \beta_{10j}*(SWBP_{ij}) + \beta_{11j}*(DISCRIM_{ij}) + \beta_{12j}*(IMMIGRCD_{ij}) + \beta_{13j}*(GENDERRC_{ij}) + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(PROAT5AM_j) + \gamma_{02}*(CLSIZE_j) + \gamma_{03}*(CREACTIV_j) + \gamma_{04}*(EDUSHORT_j) + \gamma_{05}*(TEACHBEH_j) + u_{0j}$$

$$\begin{aligned}
\beta_{1j} &= \gamma_{10} \\
\beta_{2j} &= \gamma_{20} + u_{2j} \\
\beta_{3j} &= \gamma_{30} + u_{3j} \\
\beta_{4j} &= \gamma_{40} + \gamma_{41}*(CREACTIV_j) + u_{4j} \\
\beta_{5j} &= \gamma_{50} + u_{5j} \\
\beta_{6j} &= \gamma_{60} + u_{6j} \\
\beta_{7j} &= \gamma_{70} + u_{7j} \\
\beta_{8j} &= \gamma_{80} \\
\beta_{9j} &= \gamma_{90} + u_{9j} \\
\beta_{10j} &= \gamma_{100} + u_{10j} \\
\beta_{11j} &= \gamma_{110} + u_{11j} \\
\beta_{12j} &= \gamma_{120} + \gamma_{121}*(PROAT5AM_j) + u_{12j} \\
\beta_{13j} &= \gamma_{130}
\end{aligned}$$

Mixed Model

$$\begin{aligned}
PVISCIE_{ij} &= \gamma_{00} + \gamma_{01}*PROAT5AM_j + \gamma_{02}*CLSIZE_j + \gamma_{03}*CREACTIV_j + \gamma_{04}*EDUSHORT_j + \\
&\gamma_{05}*TEACHBEH_j + \gamma_{10}*REPEAT_{ij} + \gamma_{20}*BSMJ_{ij} + \gamma_{30}*SMINS_{ij} + \gamma_{40}*ESCS_{ij} + \\
&\gamma_{41}*CREACTIV_j*ESCS_{ij} + \gamma_{50}*EMOSUPS_{ij} + \gamma_{60}*ATTLNACT_{ij} + \gamma_{70}*WORKMAST_{ij} + \\
&\gamma_{80}*GFOFAIL_{ij} + \gamma_{90}*EUDMO_{ij} + \gamma_{100}*SWBP_{ij} + \gamma_{110}*DISCRIM_{ij} + \gamma_{120}*IMMIGRCD_{ij} + \\
&\gamma_{121}*PROAT5AM_j*IMMIGRCD_{ij} + \gamma_{130}*GENDERRC_{ij} + u_{0j} + u_{2j}*BSMJ_{ij} + u_{3j}*SMINS_{ij} + \\
&u_{4j}*ESCS_{ij} + u_{5j}*EMOSUPS_{ij} + u_{6j}*ATTLNACT_{ij} + u_{7j}*WORKMAST_{ij} + u_{9j}*EUDMO_{ij} + \\
&u_{10j}*SWBP_{ij} + u_{11j}*DISCRIM_{ij} + u_{12j}*IMMIGRCD_{ij} + r_{ij}
\end{aligned}$$

Table 4.9 Results of the Final Model

Fixed Effect		Coefficient	SE	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0						
INTRCPT2,	γ_{00}	395.46	2.82	139.83	207	<.001
PROAT5AM,	γ_{01}	78.71	19.86	3.96	103	<.001
CLSIZE,	γ_{02}	1.18	0.52	2.25	207	.025
CREACTIV,	γ_{03}	8.75	3.02	2.89	207	.004
EDUSHORT,	γ_{04}	-5.66	2.77	-2.03	207	.043
TEACHBEH,	γ_{05}	6.74	2.38	2.83	207	.005
For REPEAT slope, β_1						
INTRCPT2,	γ_{10}	-37.46	11.06	-3.39	17	.003
For BSMJ slope, β_2						
INTRCPT2,	γ_{20}	0.99	0.06	16.29	125	<.001
For SMINS slope, β_3						
INTRCPT2,	γ_{30}	0.12	0.01	10.24	56	<.001
For ESCS slope, β_4						
INTRCPT2,	γ_{40}	4.02	1.41	2.86	121	.005

CREACTIV	γ_{41}	3.02	1.37	2.20	211	.029
For EMOSUPS slope, β_5						
INTRCPT2,	γ_{50}	8.59	1.68	5.11	19	<.001
For ATTLNACT slope, β_6						
INTRCPT2,	γ_{60}	-2.51	1.03	-2.43	90	.017
For WORKMAST slope, β_7						
INTRCPT2,	γ_{70}	8.46	1.04	8.15	134	<.001
For GFOFAIL slope, β_8						
INTRCPT2,	γ_{80}	-3.14	1.17	-2.69	55	.009
For EUDMO slope, β_9						
INTRCPT2,	γ_{90}	-5.94	1.28	-4.63	64	<.001
For SWBP slope, β_{10}						
INTRCPT2,	γ_{100}	-4.49	1.31	-3.43	54	.001
For DISCRIM slope, β_{11}						
INTRCPT2,	γ_{110}	-5.96	1.26	-4.73	28	<.001
For IMMIGRCD slope, β_{12}						
INTRCPT2,	γ_{120}	-16.18	4.27	-3.78	36	<.001
PROAT5AM,	γ_{121}	-107.20	47.44	-2.26	211	.025
For GENDERRC slope, β_{13}						
INTRCPT2,	γ_{130}	22.30	2.12	10.48	154	<.001
Within school variance (σ^2) = 3137.45						
Between school variance (τ) = 1204.79						

Summary of the Final Model:

There were some slight changes in the values of level-1 and level-2 coefficients as well as a new interaction being discovered, overall model did not change much from its Random Coefficient version. In this summary, more detail is put into description of interaction effects as well as the changes in the values of coefficient that showed anything more than trivial change

Among the level-2 variables only the *Index proportion of all teachers ISCED LEVEL 5A Master* (PROAT5AM), ($\gamma_{01} = 78.71$, SE = 19.86) indicated somewhat non-trivial change (dropped by 6.26 points compared to Random Effect mode.). Among level-1 variable *Index of economic, social, and cultural status* (ESCS), ($\gamma_{40} = 4.02$, SE = 1.41) increased for less than a point, and *Index Immigration status Recoded* (IMMIGRCD), ($\gamma_{120} = -16.18$, SE = 4.27) also decreased for less than a point.

Interaction effects:

Creative extra-curricular activities (CREACTIV), and *Index of economic, social, and cultural status* (ESCS), ($\gamma_{41} = 3.02$, SE = 1.37) indicates that the effect of the quantity of creative extra-curricular activities is moderated by the student's socio – economic status and vice versa. In other words, students who have a one unit increase in ESCS value as well as a one unit increase in CREATIV value will have additional 3.02 points on average more.

Index proportion of all teachers ISCED LEVEL 5A Master (PROAT5AM), and *Index Immigration status Recoded* (IMMIGRCD), ($\gamma_{421} = -107.20$, SE = 47.44) indicates that the effect of the proportion of all teacher that have master level education is moderated by the student's immigration status and vice versa. In other words, students who have a one unit increase in PROAT5AM value as well as a one unit increase in IMMIGRCD value will have additional - 107.20 points on average less.

Between school variability did not change compared to the Random Coefficient model, whereas within-school variability ($\sigma^2 = 3137.45$) trivially dropped compared to the Random Effects model ($\sigma^2 = 3140.45$).

5. DISCUSSION, CONCLUSIONS, and IMPLICATIONS

Discussion of the results begin with the limitations of the study, continuing onto the integration of findings from student-level and school-level factors with the related literature review. Conclusions follow, after which the implications for science teachers, school principals, parents and education policy makers are made. The chapter is finalized with the suggestions for further research.

5.1 Discussion of the Results

This section starts with the limitations of the study, addressing both methodological and theoretical limitations. It then continues the discussion of the findings in the context of the related literature review, first for student-level factors, and then for school-level factors.

5.1.1 Limitations of the Study

- For the missing values, Expectation-Maximization technique was used as an imputation method. Even though it is a better method than listwise deletion or Mean/Mode substitutions, it is still not as powerful as the method of Multiple imputation.
- Column-wise deletion of the variables with more than 30% of missing values was also a limitation of this study. Including deleted variables could have influenced the course of the modeling.
- The present study is associational in nature since Hierarchical Linear Modelling was used as a multilevel analysis technique. With that in line, this study does not give much information about the causes and effects of the observed relationships.
- Secondary analysis as a type of study, comes with a limitation that all reliability and validity control of test development and test administration lays outside of the control of the researcher. There is no guarantee that answers which students and school principals gave on their questionnaire are reliable, which can influence correctness of the coefficients, standard errors, and other analyses.

- Some non-normality was observed in few of the variables (kurtosis and skewness were outside of the (-2, +2) range). Multivariate normality as well as all types of residuals normality was not explicitly checked. Even though HLM is said to be robust to non-normality, having all variables univariate normal and multivariate normal, could have affected the outcomes.
- PISA's index of socio-economic and cultural status (ESCS) was the only measure of socio-economic status in this study. Including other indices, which ESCS was constructed from could have increased explained variability in science literacy scores.
- Bosnia and Herzegovina did not participate in the teacher questionnaire that was offered by PISA 2018. Deeper information about teacher characteristics and classroom processes could have significantly influence the amount of explained variability in PISA science literacy scores.
- PISA 2018 has reading as a main domain, and not science. That is important because the possible constructs that are measured more precisely concerning science literacy have not been measured in this cycle of PISA. That limited the amount of explained variability in the science literacy scores.

5.1.2 Integration of Findings from Student-level Factors with the Related Literature

Review

Out of the general student characteristics measured in PISA 2018, gender, grade level, grade repetition and time spent in science classes were significantly related to the scientific literacy scores. It has been found that gender gap in science interest does not exist in early childhood, however it increases significantly by the end of high school (Baram-Tsabari & Yarden, 2010). Some of the studies of relationship between gender and science literacy in PISA show females outperforming males (Džumhur, 2019), and some show males outperforming females (Sun et al., 2012; Üstün et al., 2020). In this study association between the gender of the students and science literacy performance was shown as significant and positive in favor of males. One of the reasons to explain that gender gap might be that science has not been traditionally viewed as a female interest.

PISA defines their target population based on age, including only 15-year-olds enrolled in at least 7th grade of primary education, which explains why age was not statistically significant

predictor of science literacy. There's proof however, of students in higher grades performing academically better than those in lower grades, even if they are of the same age (Marks, 2006). In the present study, higher grade level of the students indicated higher science literacy scores. Most (more than 70%) of the student sample from Bosnia and Herzegovina in PISA 2018 was from the 1st grade of secondary education, which means that it was to be expected that those students perform better than the students in the final year of primary education. However, grade ceased to be a significant variable when controlling for the other student- and school-level factors. An explanation for that besides the same age of students is that other variables explained variability in the population, even if they were in different grades. According to the present study, the students that repeat a grade on average performed worse than the students that did not. This result is not surprising because the students that fail a year have a higher chance to be stigmatized (Gottfredson et al., 1994).

Time spent learning science was found to be significantly and positively related to science literacy scores. An issue with this variable is primarily related to the fact that it measures the number of minutes spent in science related classes, as reported by the students. It does not measure, as one would expect, amount of time the student invests in learning science on its own, which was proven to influence science achievement (Kidron & Lindsay, 2014).

The environment in which learning is happening was shown to be an important factor in science achievement (Ho, 2010; Juan & Visser, 2017). In this study, learning environment factors such as: parents emotional support, immigration status, discriminating school climate, and socio-economic and cultural status showed significant statistical relationships with science literacy. The students that felt emotionally supported by their parents scored better on PISA's scientific literacy. In line with studies that confirm positive effect of family support on science achievement (Ho, 2010; Mashile, 2001), this was not a surprising result. Not only it is shown that students that have immigration background are disadvantaged compared to native students, but also that those disadvantages become more pronounced when schools in which they go to have higher number of teachers with Master diploma (Hsu, 2007). Discriminating school climate was found to be significantly and negative predictor of science literacy. However, student experience of being bullied and student experience of belonging to school did not show any significant effect on scientific literacy. The socio-economic and cultural status of the student was found to be significantly and positively correlated with science literacy scores. This was to be expected, since

there was a lot of confirmation that socio-economic status positively influences school achievement (Sirin, 2005) and science achievement (Özer & Anil, 2011; Üstün et al., 2020).

Individual student characteristics such as: motivation to master tasks and attitude towards learning positively predict science literacy which was to be expected from the literature (Atwater et al., 1995; Ceylan & Berberoglu, 2007; Singh et al., 2002). On the other hand: fear of failure, meaning in life and positive feelings negatively predict science literacy. A negative effect of fear of failure which is just a measure of anxiety on science literacy was to be expected (Pitsia et al., 2017). However, positive feeling and meaning in life negative effect was not expected. One explanation could be that 15-year-olds might not have a clear meaning in life yet, due to still being quite young. In addition, the fact that science can be very challenging and it requires cognitive load which might not be very comfortable, negative relationship with positive feelings might make more sense. Self-efficacy (resilience) and competitiveness were not found to be significant in predicting science literacy, which was surprising. The reason for that might be that self-efficacy in the context of science achievement wasn't measured as in other studies which found significant relationships (Sun et al., 2012).

5.1.3 Integration of Findings from School-level Factors with the Related Literature Review

General school context as measured from the perspective of the school principal was found to be an important factor as well in this study. In the schools in which there are more creative-extracurricular activities, the students perform better in science literacy. Extra-curricular activities are related to students' self-esteem, social skills and academic achievement (Fujita, 2006; Ponter, 1999). In addition, interaction effect was found between socio-economic and cultural status of the student and the quantity of school's extra-curricular activities in the context of science literacy. That could indicate that the students of higher SES benefit more from the extra-curricular activities and vice versa. In the schools in which the shortage of educational material in schools is more pronounced, students' science literacy scores are on average lower. Condition of school buildings has been found to be significant indicator of science achievement, as shown with South African students in TIMSS 2012 (Juan & Visser, 2017). Since a lot of the science classes require special equipment to explore the natural phenomena, the negative effect of their lack, makes sense (Altun & Kalkan, 2021). Proportion of available computers that are connected to the Internet was not

found to be statistically significant as a measure that affects science literacy scores, which was surprising because some other showed significant positive effect. (Üstün et al., 2020)

Out of the teachers' characteristics in the schools in which there is more teachers with Masters diploma, students perform better in science literacy. The more knowledge teachers have, the more they can teach (Zhang, 2008). However, surprisingly in the schools in which there is more teacher hindering learning, students perform better in science literacy. Again, since science is challenging, teacher that are stricter might get better results from their students.

Class Size was the only measure of class context according to (Helmke, 2012) and it was found to be significantly and positively correlated with science literacy scores of the students. Even though there are some findings that smaller class sizes can improve study outcomes (Dynarski et al., 2013), while others are inconclusive (Woessmann & West, 2006), there is also proof of bigger class sizes having beneficial effect (Džumhur, 2019).

5.2 Conclusions

The present study aimed to explore which student- and school-level factors, as well as their interactions have significant relationships with Bosnia and Herzegovina students' scientific literacy scores as measured in PISA 2018. Based on the correlational study, using Hierarchical Linear Modeling, we can conclude that we explain factors affecting the students' scientific literacy scores with a revised Supply-Use model shown in Figure 5.1.

Table 5.1 Revised Supply-Use model

Student – level variables (level-1)
<i>General student characteristics:</i>
Student gender
Grade repetition
Minutes of Student's time spend in science classes per week
<i>Learning environment</i>
Discriminating school climate
Parents' emotional support perceived by student
Socio-economic and cultural status
Immigration status
<i>Individual student characteristics:</i>
Motivation to master tasks
General fear of failure
Eudaemonia: meaning in life
Subjective well-being: Positive feelings
Student's expected occupational status
Attitude towards school: learning activities
School – level variables (level-2)
<i>General school context:</i>
Quantity of creative extra-curricular activities
Shortage of educational material
<i>Teacher characteristics</i>
Proportion of all teachers with Master degree
Teacher behavior hindering learning
<i>Classroom context:</i>
Class Size
Interaction effects
Immigration status of the student and the proportion of all teachers with Master degree at school
Socio-economic and cultural status and the quantity of creative extra-curricular activities at school

5.3 Implications

According to the results of the present study, we have some implications for science teachers, principals of schools, parents of students, and policy makers as follows:

- Science teachers:
 - should work towards raising awareness of the gender gap in science interest and work towards supporting females in their pursuit of science.
 - should show special care and sensitivity towards vulnerable student groups (such as student that repeated a grade or have immigrant status) and be careful not to discriminate anyone based on their cultural background.
- Principals of schools:
 - should support and advocate for the increase of extra-curricular activities, as well as make sure that those activities are available to all students, no matter their socio-economic background.
 - should make sure that they have good communication with their teachers, as to better understand and be able to evaluate teacher behaviors that hinder learning as well as teachers needs for the educational materials (e.g., textbooks, IT equipment, library, or laboratory material).
- Parents of students:
 - should find ways to support their children emotionally, even if they are not able to support them in other ways such as socio-economically or academically.
- Educational policy makers:
 - should view participation in upcoming cycles of PISA as an imperative to be able to systematically approach bettering of science education in Bosnia and Herzegovina.
 - should not miss an opportunity for Bosnia and Herzegovina to participate in upcoming PISA's Teacher Questionnaires.

5.3.1 Suggestions for Further Research

In line with the limitations of this study and the literature review, some suggestions for further research are described below.

Investigating the effects of various variables which PISA's ESCS index is made of, such as: home possessions, education level of mother and father, number of educational materials in house etc. That could increase amount of explained variability in science literacy due to socio-economic status.

Taking into consideration between-country variability of science literacy scores in PISA 2018. For example, relationship between student and school factors and science literacy in PISA 2018 across the countries of former Republic of Yugoslavia could be interesting to research because all those countries belonged to a same educational system at some point in time.

Instead of Hierarchical Linear Modelling, Multilevel Structural Equation Modeling could have been used as a different approach to modelling PISA 2018 scientific literacy data for Bosnia and Herzegovina.

REFERENCES

- Adams, R., & Wu, M. (2002). *PISA 2000*. Technical report.
- Altun, A., & Kalkan, Ö. K. (2021). Cross-national study on students and school factors affecting science literacy. *Educational Studies*, 47(4), 403–421.
- Atwater, M. M., Wiggins, J., & Gardner, C. M. (1995). A study of urban middle school students with high and low attitudes toward science. *Journal of Research in Science Teaching*, 32(6), 665–677.
- Baram-Tsabari, A., & Yarden, A. (2010). QUANTIFYING THE GENDER GAP IN SCIENCE INTERESTS. *International Journal of Science and Mathematics Education 2010* 9:3, 9(3), 523–550. <https://doi.org/10.1007/S10763-010-9194-7>
- Barrett, A. M., Chawla-Duggan, R., Lowe, J., Nickel, J., & Ukpo, E. (2006). The concept of quality in education: A review of the ‘international’ literature on the concept of quality in education. *England: EdQual*.
- Baumert, J., Blum, W., & Neubrand, M. (2004). Drawing the lessons from PISA 2000. *Zeitschrift Für Erziehungswissenschaft*, 7(3), 143–157.
- Bloom, B. S., Block, J. H., Airasian, P. W., & Carroll, J. B. (1971). *Mastery learning: Theory and practice*. Holt, Rinehart and Winston,.
- Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction*, 21(1), 95–108.
- Carroll, J. (1963). A model of school learning. *Teachers College Record*, 64(8), 723.
- Ceylan, E., & Abaci, S. (2013). Differences between Turkey and Finland based on Eight Latent Variables in PISA 2006. *International Online Journal of Educational Sciences*, 5(1).

- Ceylan, E., & Berberoglu, G. (2007). Factors Related With Students' Science Achievement: A Modeling Study. *Egitim ve Bilim*, 32(144), 36.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.
- DeBoer, G. E. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 37(6), 582–601.
- Dewey, J. (1923). *Democracy and education: An introduction to the philosophy of education*. macmillan.
- Dynarski, S., Hyman, J., & Schanzenbach, D. W. (2013). Experimental evidence on the effect of childhood investments on postsecondary attainment and degree completion. *Journal of Policy Analysis and Management*, 32(4), 692–717.
- Džumhur, Ž. (2019). *PISA 2018 report for Bosnia and Herzegovina | APOSO*. Agencija Za Predškolsko, Osnovno i Srednje Obrazovanje. <https://doi.org/ISBN 978-9958-572-13-5>
- Finn, J. D. (1989). Withdrawing From School. *Review of Educational Research*, 59(2), 117–142. <https://doi.org/10.3102/00346543059002117>
- Fujita, K. (2006). The effects of extracurricular activities on the academic performance of junior high students. *Undergraduate Research Journal for the Human Sciences*, 5(1), 1–16.
- Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction*. Longman Publishing.
- Ganzeboom, H. B. G., & Treiman, D. J. (2003). Three internationally standardised measures for comparative research on occupational status. In *Advances in cross-national comparison* (pp. 159–193). Springer.
- Glass, G. V., & Smith, M. L. (1979). Meta-Analysis of Research on Class Size and Achievement.

Educational Evaluation and Policy Analysis, 1(1), 2–16.

<https://doi.org/10.3102/01623737001001002>

Gottfredson, D. C., Fink, C. M., & Graham, N. (1994). Grade Retention and Problem Behavior.

American Educational Research Journal, 31(4), 761–784.

<https://doi.org/10.3102/00028312031004761>

Gungor, A., Eryilmaz, A., & Fakıoglu, T. (2007). The relationship of freshmen's physics achievement and their related affective characteristics. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 44(8), 1036–1056.

Helmke, A. (2012). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts. Franz Emanuel Weinert gewidmet. 3. Aufl.* (F. E. Weinert (ed.)).

Kallmeyer u.a.

Helmke, A., & Lenske, G. (2013). Unterrichtsdiagnostik als Voraussetzung für Unterrichtsentwicklung. *Beiträge Zur Lehrerinnen- Und Lehrerbildung*, 31(2).

Ho, E. S. C. (2010). Family influences on science learning among Hong Kong adolescents: What we learned from pisa. *International Journal of Science and Mathematics Education*, 8(3), 409–428. <https://doi.org/10.1007/s10763-010-9198-3>

Hopfenbeck, T. N., Lenkeit, J., El Masri, Y., Cantrell, K., Ryan, J., & Baird, J.-A. (2018). Lessons learned from PISA: A systematic review of peer-reviewed articles on the programme for international student assessment. *Scandinavian Journal of Educational Research*, 62(3), 333–353.

Hox, J. (1998). *Multilevel Modeling: When and Why. 1965*, 147–154. https://doi.org/10.1007/978-3-642-72087-1_17

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and*

- applications*. Routledge.
- Hsu, J.-C. (2007). *Comparing the relationships between mathematics achievement and student characteristics in Canada and Hong Kong through HLM*.
- Huitt, W., & McIlrath, D. (1995). The teaching-learning process: A discussion of models. *Educational Psychology Interactive*, 45–49.
- Hurd, P. D. (1958). Science literacy: Its meaning for American schools. *Educational Leadership*, 16(1), 13–16.
- Husremović, D., & Đapo, N. (2013). *Sekundarne analize podataka –prediktora uspjeha na testovima iz matematike i bosanskog, hrvatskog i srpskog jezika*. <https://aposo.gov.ba/bs/benchmarking/>
- IBM. (2019). *IBM SPSS Statistics 26 Documentation*. <https://www.ibm.com/support/pages/ibm-spss-statistics-26-documentation#de>
- Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics*, 86(1), 226–244. <https://doi.org/10.1162/003465304323023778>
- Juan, A., & Visser, M. (2017). Home and school environmental determinants of science achievement of South African students. *South African Journal of Education*, 37(1), 1–10.
- Karasar, N. (2012). Bilimsel araştırma yöntemi [Scientific research method]. *Ankara: Nobel Yayınları*.
- Kidron, Y., & Lindsay, J. (2014). *The effects of increased learning time on student academic and nonacademic outcomes: Findings from a meta-analytic review*.
- Kircher, E., Girwidz, R., & Häußler, P. (2009). *Physikdidaktik: Theorie und Praxis*. Springer-Verlag.
- Kremer, B. K., & Walberg, H. J. (1981). *A Synthesis of Social and Psychological Influences on*

- Science Learning. *Science Education*, 65(1), 11–23.
- Laugksch, R. C. (2000). Scientific literacy: A conceptual overview. *Science Education*, 84(1), 71–94.
- Lenzen, D., Baumert, J., Watermann, R., & Trautwein, U. (2004). *PISA und die Konsequenzen für die erziehungswissenschaftliche Forschung* (Vol. 3). Springer-Verlag.
- Luu, K., & Freeman, J. G. (2011). An analysis of the relationship between information and communication technology (ICT) and scientific literacy in Canada and Australia. *Computers & Education*, 56(4), 1072–1082. <https://doi.org/10.1016/j.compedu.2010.11.008>
- Marks, G. N. (2006). Are between- and within-school differences in student performance largely due to socio-economic background? Evidence from 30 countries. *Educational Research*, 48(1), 21–40. <https://doi.org/10.1080/00131880500498396>
- Mashile, E. O. (2001). Science achievement determinants: factorial structure of family variables. *South African Journal of Education*, 21(4), 336–338.
- Mattern, N., & Schau, C. (2002). Gender differences in science attitude-achievement relationships over time among white middle-school students. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 39(4), 324–340.
- McFarlane, D. A. (2013). Understanding the challenges of science education in the 21st century: New opportunities for scientific literacy. *International Letters of Social and Humanistic Sciences*, 04, 35–44.
- Mettas, A., Karmiotis, I., & Christoforou, P. (2006). Relationship between students' self-beliefs and attitudes on science achievements in Cyprus: findings from The Third International Mathematics and Science Study (TIMSS). *Eurasia Journal of Mathematics, Science and Technology Education*, 2(1), 41–52.

- Nasr, A. R., & Soltani, A. K. (2011). Attitude towards biology and its effects on student's achievement. *International Journal of Biology*, 3(4), 100.
- OECD. (2009a). *PISA Data Analysis Manual: SPSS® Second Edition*. Author Paris, France.
- OECD. (2009b). *PISA Data Analysis Manual: SPSS® Second Edition*. Author Paris, France.
<https://doi.org/10.1787/9789264056275-en>
- OECD. (2016a). PISA 2015 Assessment and Analytical Framework PISA 2015 Assessment and Analytical Framework. In *OECD Publishing*. OECD publishing. <https://www.oecd-ilibrary.org/docserver/9789264190511-en.pdf?expires=1569847112&id=id&accname=guest&checksum=08AEA3FD9105123D4555A383BD097B5E>
- OECD. (2016b). *PISA 2015 results (volume I): Excellence and equity in education*. OECD Publishing.
- OECD. (2019a). *Chapter 16 - Scaling procedures and construct validation of context questionnaire data*. PISA 2018 Technical Report. <https://www.oecd.org/pisa/data/pisa2018technicalreport/>
- OECD. (2019b). *Chapter 2 - Test Design*. PISA 2018 Technical Report.
<https://www.oecd.org/pisa/data/pisa2018technicalreport/>
- OECD. (2019c). Chapter 3 - Context Questionnaire Development. *PISA 2018 Technical Report*.
<https://www.oecd.org/pisa/data/pisa2018technicalreport/>
- OECD. (2019d). Chapter 8 - Survey Weighting and the Calculation of Sampling Variance. *PISA 2018 Technical Report*.
- OECD. (2019e). *How to prepare and analyse the PISA database*.
<https://www.oecd.org/pisa/data/httpoecdorgpisadatabase-instructions.htm>
- OECD. (2019f). *PISA 2018 Results (Volume I): Vol. I*. <https://doi.org/10.1787/5f07c754-en>
- OECD. (2020). *Discover the OECD TOGETHER WE CREATE BETTER POLICIES FOR BETTER*

LIVES.

- Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, 25(9), 1049–1079.
- Ozel, M., Caglak, S., & Erdogan, M. (2013). Are affective factors a good predictor of science achievement? Examining the role of affective factors based on PISA 2006. *Learning and Individual Differences*, 24, 73–82.
- Özer, Y., & Anil, D. (2011). ÖĞRENCİLERİN Fen Ve Matematik Başarılarını Etkileyen Faktörlerin Yapısal Eşitlik Modeli ile İncelenmesi. *Hacettepe Eğitim Dergisi*, 41(41), 313–324.
- Perera, L. D. H. (2014). Parents' Attitudes Towards Science and their Children's Science Achievement. *International Journal of Science Education*, 36(18), 3021–3041. <https://doi.org/10.1080/09500693.2014.949900>
- Pitsia, V., Biggart, A., & Karakolidis, A. (2017). The role of students' self-beliefs, motivation and attitudes in predicting mathematics achievement: A multilevel analysis of the Programme for International Student Assessment data. *Learning and Individual Differences*, 55, 163–173. <https://doi.org/10.1016/j.lindif.2017.03.014>
- Ponter, J. R. (1999). Academic achievement and the need for a comprehensive, developmental music curriculum. *NASSP Bulletin*, 83(604), 108–114.
- Proctor, C. P. (1984). Teacher Expectations: A Model for School Improvement. *The Elementary School Journal*, 84(4), 469–481. <https://doi.org/10.1086/461377>
- Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods*. https://books.google.com/books?hl=en&lr=&id=uyCV0CNGDLQC&oi=fnd&pg=PR17&dq=An+introduction+to+hierarchical+linear+modeling&ots=qC1MTn_ZF&sig=zGVyYkxco4sJT_p081H4ao-3Xo8

- Rennie, L. J., & Punch, K. F. (1991). The relationship between affect and achievement in science. *Journal of Research in Science Teaching*, 28(2), 193–209.
- Rubba, P. A., & Andersen, H. O. (1978). Development of an instrument to assess secondary school students understanding of the nature of scientific knowledge. *Science Education*, 62(4), 449–458.
- Seidel, T. (2015). Performance assessment and professional development in university teaching. In *Incentives and Performance: Governance of Research Organizations* (pp. 465–477). Springer International Publishing. https://doi.org/10.1007/978-3-319-09785-5_28
- Simpson, R. D., & Oliver, S. J. (1990). A summary of major influences on attitude toward and achievement in science among adolescent students. *Science Education*, 74(1), 1–18. <https://doi.org/10.1002/sce.3730740102>
- Singh, K., Granville, M., & Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interest, and academic engagement. *The Journal of Educational Research*, 95(6), 323–332.
- Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, 75(3), 417–453. <https://doi.org/10.3102/00346543075003417>
- Sun, L., Bradley, K. D., & Akers, K. (2012). A Multilevel Modelling Approach to Investigating Factors Impacting Science Achievement for Secondary School Students: PISA Hong Kong Sample. *International Journal of Science Education*, 34(14), 2107–2125. <https://doi.org/10.1080/09500693.2012.708063>
- Tabachnick, B. G. 1936-, & Fidell, L. S. 1942-. (2014). *Using multivariate statistics*. Pearson.
- Uguroglu, M. E., & Walberg, H. J. (1979). Motivation and achievement: A quantitative synthesis. *American Educational Research Journal*, 16(4), 375–389.

- Usak, M., Prokop, P., Ozden, M., Ozel, M., Bilen, K., & Erdogan, M. (2009). *TURKISH UNIVERSITY STUDENTS' ATTITUDES TOWARD BIOLOGY: THE EFFECTS OF GENDER AND ENROLMENT IN BIOLOGY CLASSES*.
- Üstün, U., Özdemir, E., Cansiz, M., & Cansiz, N. (2020). What are the factors affecting turkish students' science literacy? A hierarchical linear modelling study using pisa 2015 data. *Hacettepe Egitim Dergisi*, 35(3), 720–732. <https://doi.org/10.16986/HUJE.2019050786>
- Ward, M. (2018). *PISA for development: Results in focus*.
- Watters, J. J., & Ginns, I. S. (2000). Developing motivation to teach elementary science: Effect of collaborative and authentic learning practices in preservice education. *Journal of Science Teacher Education*, 11(4), 301–321. <https://doi.org/10.1023/A:1009429131064>
- Willson, V. L., Ackerman, C., & Malave, C. (2000). Cross-time attitudes, concept formation, and achievement in college freshman physics. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 37(10), 1112–1120.
- Woessmann, L., & West, M. (2006). Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS. *European Economic Review*, 50(3), 695–736.
- Zhang, D. (2008). *The effect of teacher education level, teaching experience, and teaching behaviors on student science achievement*. Utah State University.

APPENDICES

APPENDIX A

A.1 Descriptive Statistics

A. 1.1 Histograms for Level-1 predictors

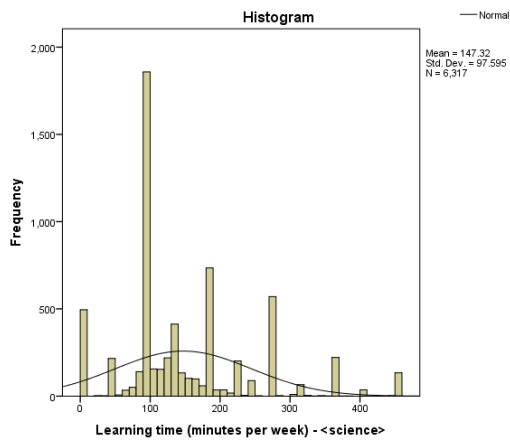


Figure A.1.1 Histogram of SMINS

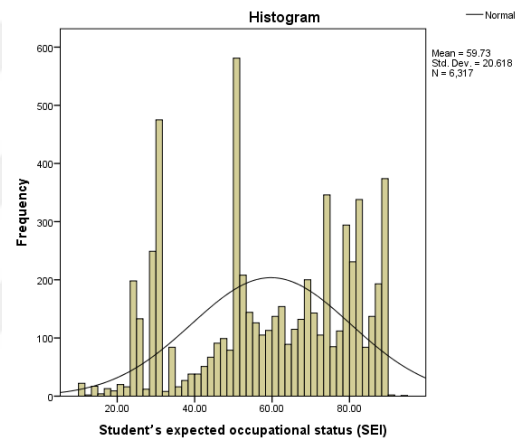


Figure A.1.2 Histogram of BSMJ

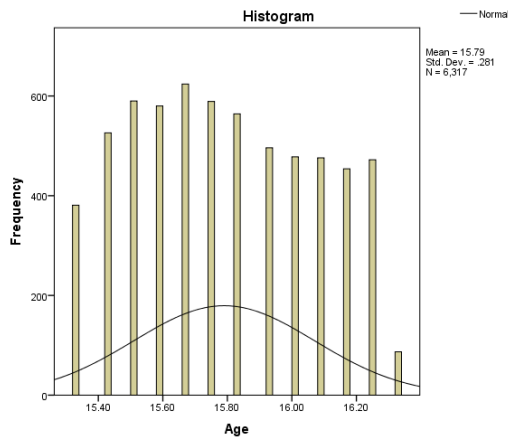


Figure A.1.3 Histogram of AGE

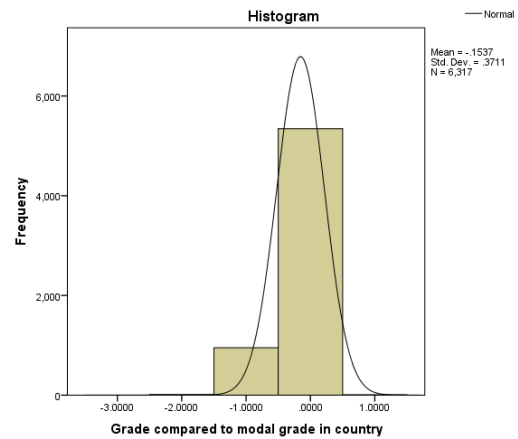


Figure A.1.4 Histogram of GRADE

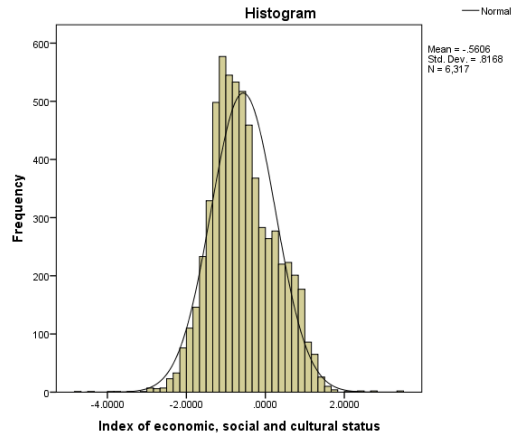


Figure A.1.5 Histogram of ESCS

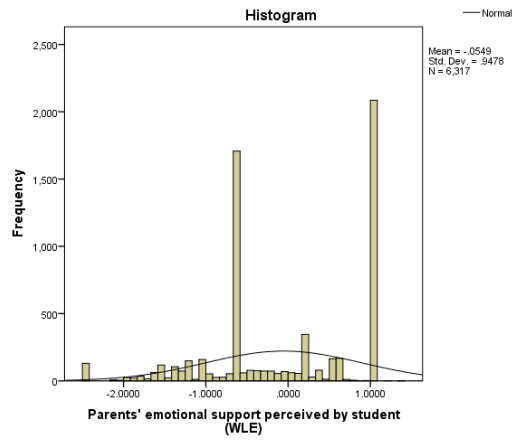


Figure A.1.6 Histogram of EMOSUPS

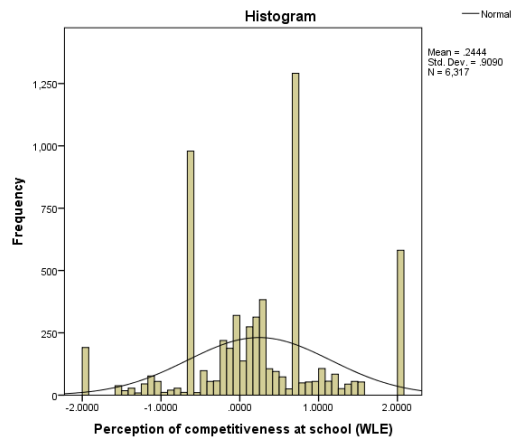


Figure A.1.7 Histogram of PERCOMP

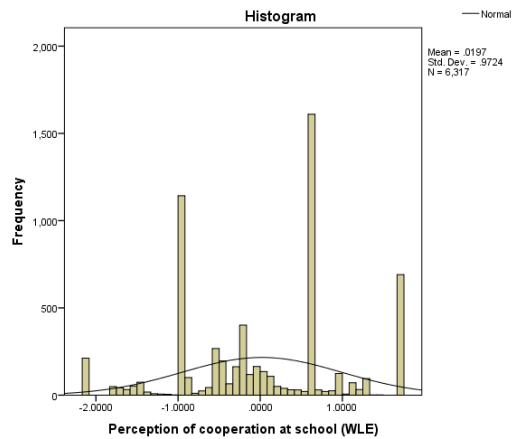


Figure A.1.8 Histogram of PERCOOP

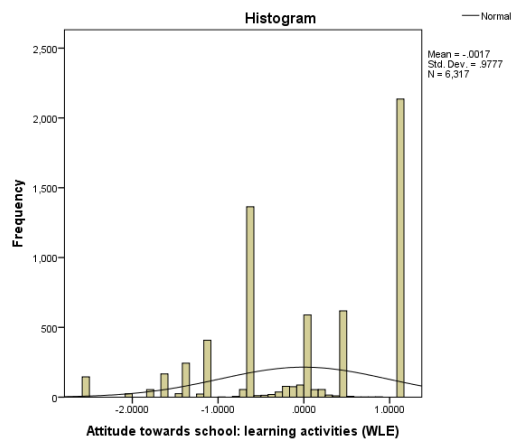


Figure A.1.9 Histogram of ATTLNACT

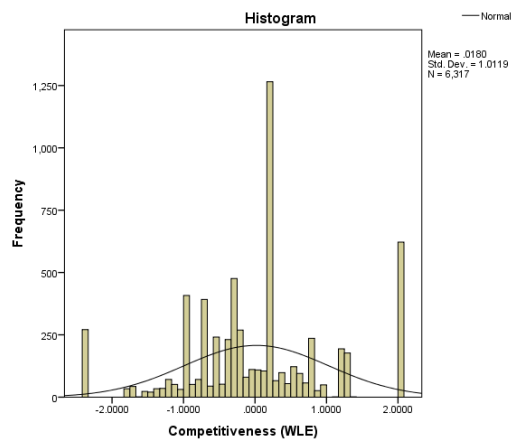


Figure A.1.10 Histogram of COMPETE

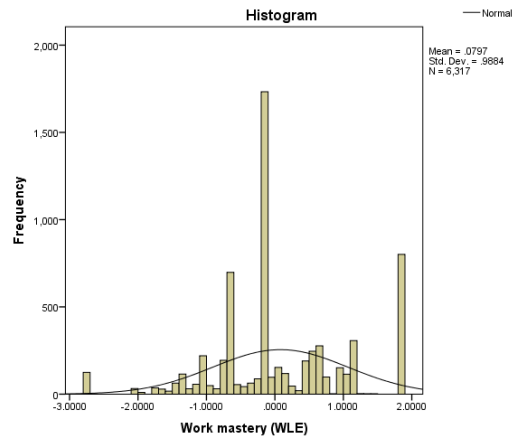


Figure A.1.11 Histogram of WORKMAST

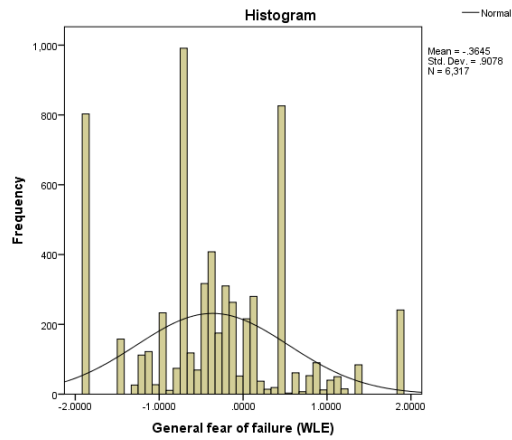


Figure A.1.12 Histogram of GFOFAIL

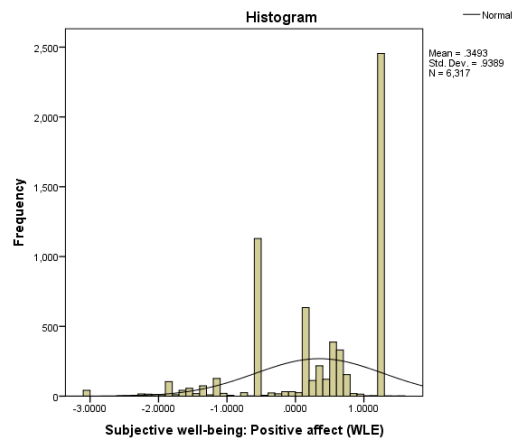


Figure A.1.13 Histogram of SWBP

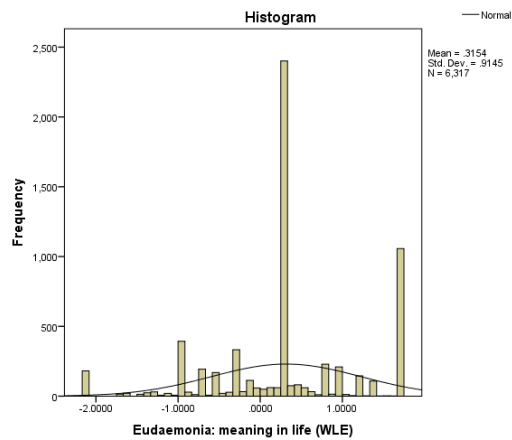


Figure A.1.14 Histogram of EUDMO

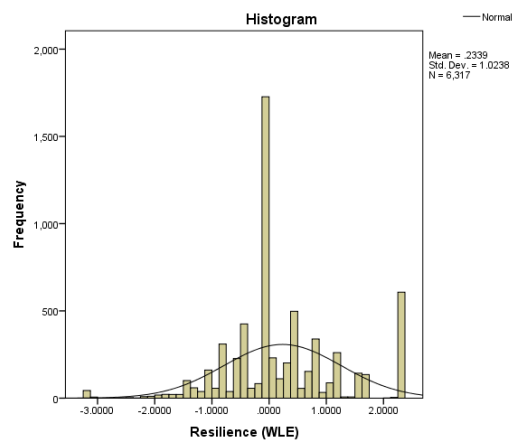


Figure A.1.15 Histogram of RESILIENCE

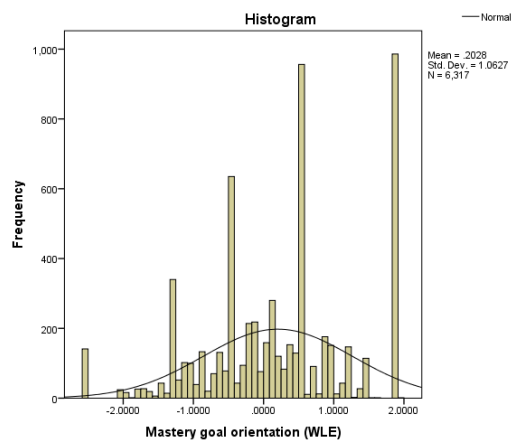


Figure A.1.16 Histogram of MASTGOAL

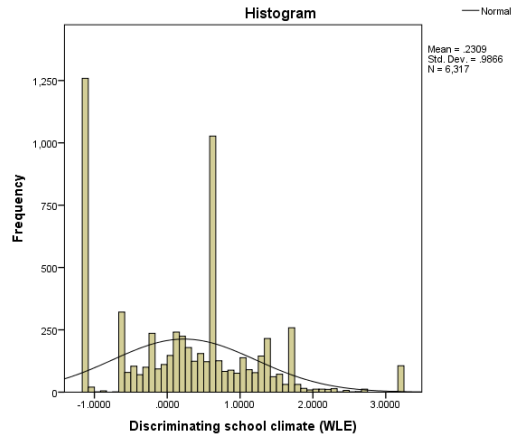


Figure A.1.17 Histogram of DISCRIM

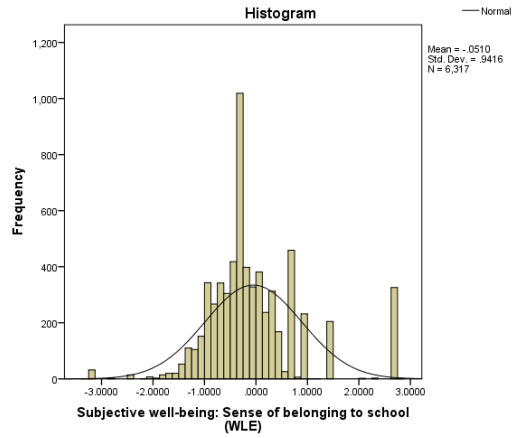


Figure A.1.18 Histogram of BELONG

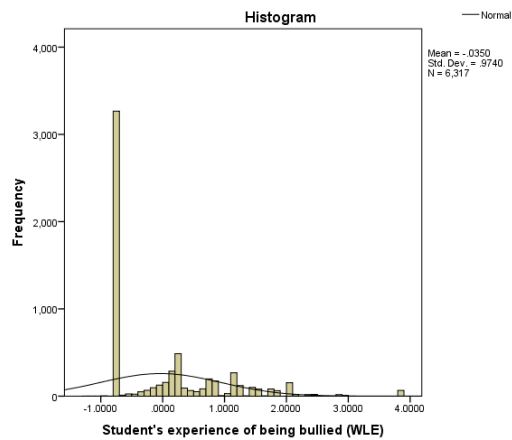


Figure A.1.19 Histogram of BEINGBULL

A. 1.2 Histograms for Level – 2 predictors

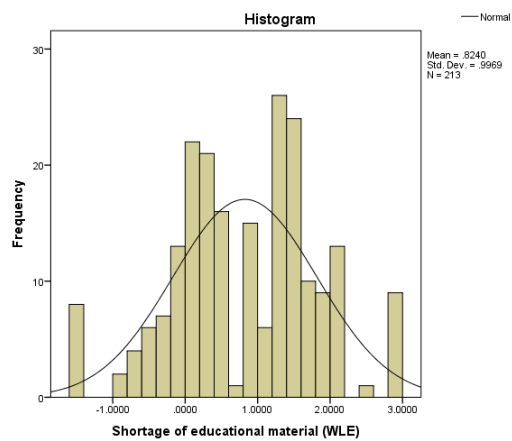


Figure A.2.1 Histogram of EDUSHORT

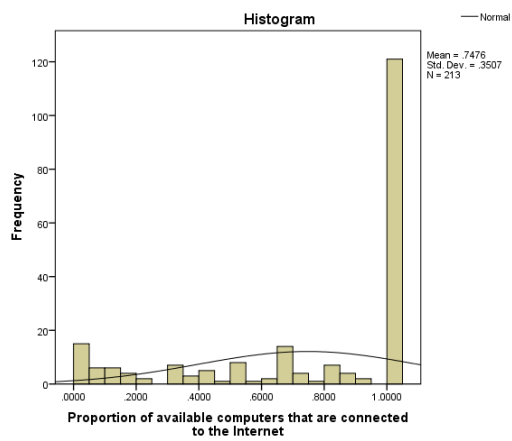


Figure A.2.2 Histogram of RATCMP2

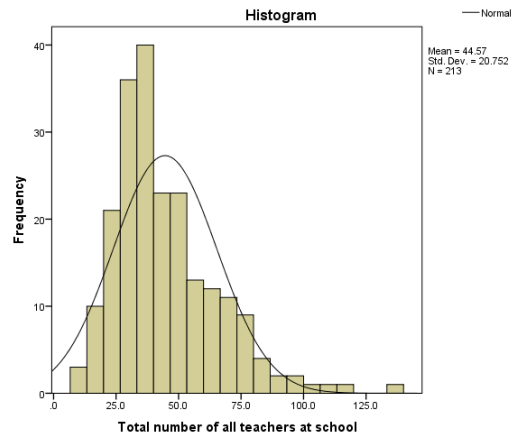


Figure A.2.3 Histogram of TOTAT

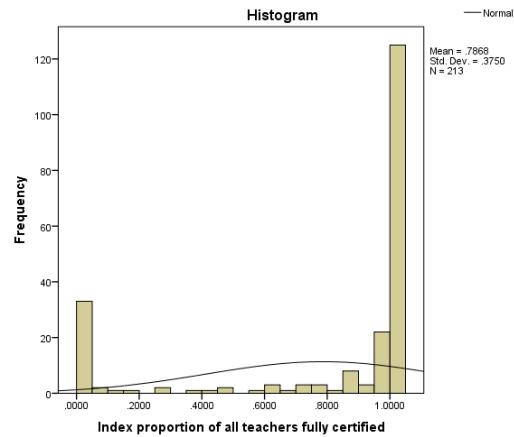


Figure A.2.4 Histogram of PROATCE

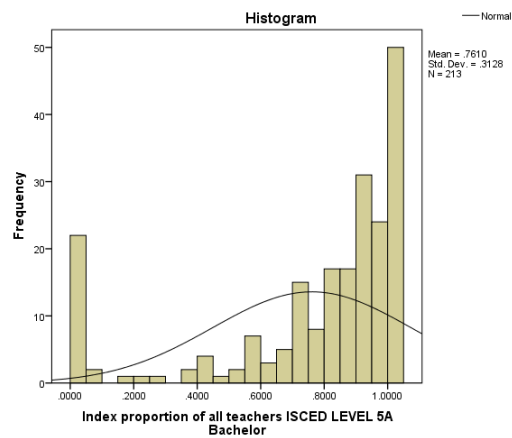


Figure A.2.5 Histogram of PROAT5AB

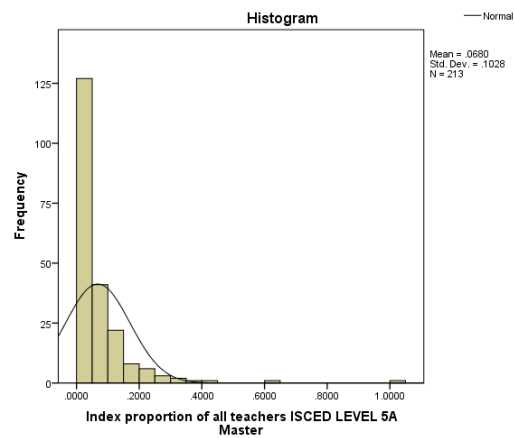


Figure A.2.6 Histogram of PROAT5AM

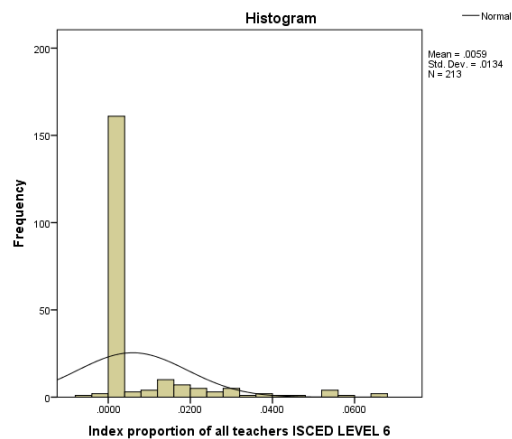


Figure A.2.7 Histogram of PROAT6

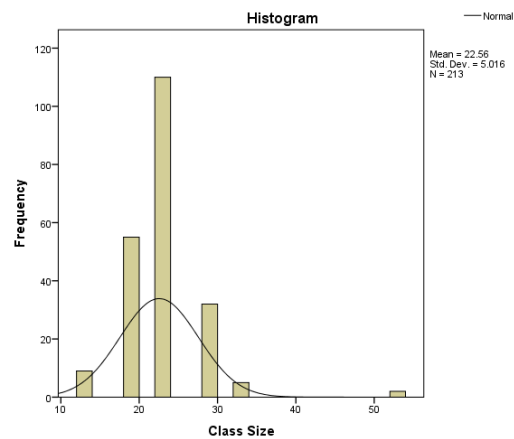


Figure A.2.8 Histogram of CLSIZE

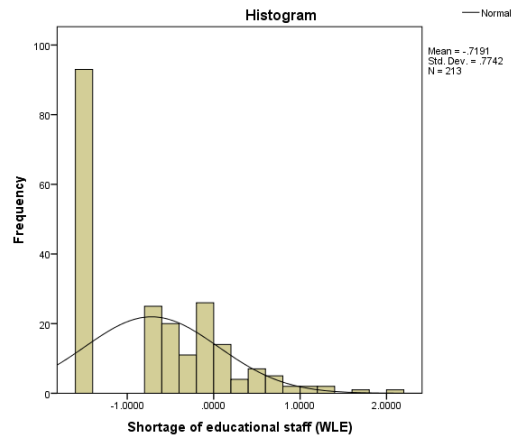


Figure A.2.9 Histogram of STAFFSHORT

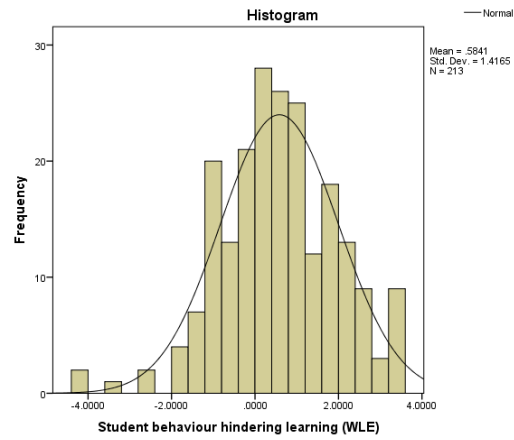


Figure A.2.10 Histogram of STUBEHA

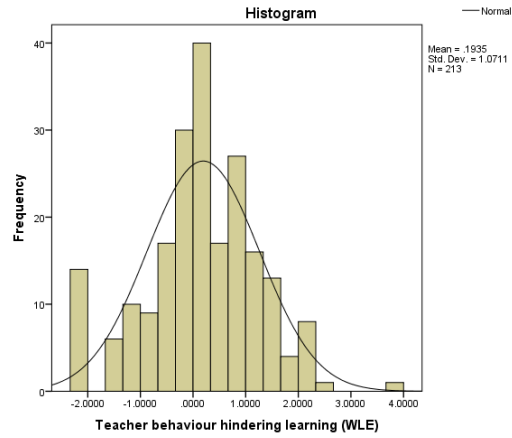


Figure A.2.11 Histogram of TEACHBEH

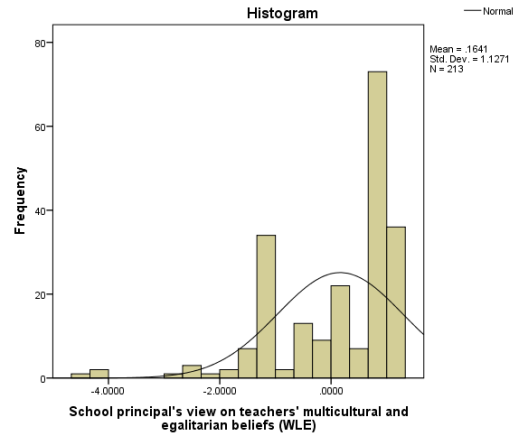


Figure A.2.12 Histogram of SCMCEG

A 2.3 Descriptive statistics (before Expectation Maximization)

Table A. 1 Descriptive statistics (level-1 data) before Expectation-Maximization missing data replacement

LEVEL-1								
VARIABLES	N	Mean	Median	Min	Max	SD	Kurtosis	Skewness
GENDERRCD	6317	A dummy variable (Student (Standardized) Gender Recoded) with following values: 0 = "Female" (f = 3082), 0 = "Male" (f = 3235).						
IMMIGRCD	6068	A dummy ordinal variable (Index Immigration Status Recoded) with following values: 0 = "Native student" (f = 5916), 1 = "Second-generation student" (f = 83), 1 = "First-generation student" (f = 87).						
REPEAT	6152	A dummy variable (Grade Repetition) with following values: 0 = "Did not repeat a grade" (f = 6042), 1 = "Repeated a grade" (f = 110).						
GRADE	6317	-0.15	0.00	-3.00	1.00	0.37	3.03	-2.01
AGE	6317	15.79	15.75	15.33	16.33	0.28	-1.10	0.12
BSMJ	5147	60.33	65.12	11.56	88.96	22.07	-1.22	-0.39
SMINS	5048	151.63	90.00	.00	450.00	107.45	0.21	0.84
ESCS	6216	-0.56	-0.66	-4.79	3.42	0.82	0.06	0.28
EMOSUPS	5440	-0.02	0.21	-2.45	1.03	1.00	-1.02	-0.36
PERCOMP	5289	0.26	0.29	-1.99	2.04	0.98	-0.26	0.00
PERCOOP	5267	0.03	0.60	-2.14	1.68	1.05	-0.89	-0.12
ATTLNACT	5852	0.00	0.03	-2.54	1.08	1.01	-0.85	-0.44
COMPETE	5822	0.02	0.20	-2.35	2.01	1.05	-0.01	0.07
WORKMAST	5636	0.11	-0.10	-2.74	1.82	1.03	0.02	-0.11
GFOFAIL	5721	-0.37	-0.43	-1.89	1.89	0.95	-0.19	0.24
EUDMO	5754	0.31	0.26	-2.15	1.74	0.95	-0.03	-0.27
SWBP	5691	0.35	0.65	-3.07	1.24	0.98	0.10	-0.91
RESILIENCE	5699	0.25	-0.06	-3.17	2.37	1.07	0.39	0.27
MASTGOAL	5729	0.21	0.35	-2.53	1.85	1.11	-0.52	-0.21
DISCRIM	5237	0.21	0.32	-1.15	3.18	1.07	-0.42	0.37
BELONG	5721	-0.04	-0.32	-3.26	2.78	0.98	1.98	0.96
BEINGBULLIED	5178	-0.06	-0.78	-0.78	3.86	1.06	1.17	1.35
PV1SCIE	6317	398.04	395.50	152.33	663.90	75.59	-0.22	0.16
PV2SCIE	6317	398.47	395.53	146.99	657.02	75.61	-0.16	0.16
PV3SCIE	6317	397.50	395.14	147.12	665.04	76.19	-0.16	0.13
PV4SCIE	6317	398.76	396.06	161.19	682.09	76.92	-0.06	0.16

PV5SCIE	6317	397.54	393.98	131.96	703.52	76.33	-0.08	0.21
PV6SCIE	6317	398.94	395.37	140.78	681.24	76.76	-0.20	0.16
PV7SCIE	6317	397.20	394.19	153.43	666.54	76.26	-0.23	0.12
PV8SCIE	6317	398.64	395.31	137.06	688.10	76.28	-0.22	0.14
PV9SCIE	6317	398.19	396.76	143.08	684.23	75.65	-0.12	0.13
PV10SCIE	6317	398.13	395.22	161.47	676.09	76.88	-0.17	0.21

Note: Descriptive statistics given in this table are given for N = 6317 students. Without missing data replacement, but with the exclusion of invalid SMINS scores.

Table A. 2 Descriptive statistics (level-2 data) before Expectation-Maximization missing data replacement

LEVEL-2								
VARIABLES		Mean	Median	Min	Max	SD	Kurtosis	Skewness
RATCMP2	211	0.75	1.00	0.00	1.00	0.35	-0.42	-1.06
TOTAT	212	44.64	39.00	11.00	136.50	20.78	1.97	1.20
PROATCE	199	0.79	1.00	0.00	1.00	0.39	0.18	-1.44
PROAT5AB	203	0.76	0.91	0.00	1.00	0.32	1.05	-1.54
PROAT5AM	211	0.07	0.04	0.00	1.00	0.10	35.89	4.94
PROAT6	207	0.01	0.00	0.00	0.07	0.01	6.61	2.59
CLSIZE	212	22.55	23.00	13.00	53.00	5.03	11.50	2.05
CREACTIV	A dummy (ordinal) variable (Creative extra-curricular activities (Sum)) as explained with following values: 0 = “No extra – curricular activities”, (f = 8), 1 = “1 type of extra – curricular activities” (f = 26), 2 = “2 types of of extra – curricular activities” (f = 59), 3 = “3 types of extra – curricular activities” (f = 120).							
EDUSHORT	213	0.82	0.83	-1.42	2.96	1.00	-0.32	-0.02
STAFFSHORT	213	-0.72	-0.78	-1.47	2.09	0.77	0.14	0.79
STUBEHA	212	0.59	0.56	-4.03	3.58	1.42	0.52	-0.20
TEACHBEHA	212	0.20	0.23	-2.04	3.79	1.07	0.16	-0.12
SCMCEG	212	0.16	0.90	-4.59	1.29	1.13	1.90	-1.30

Note: Descriptive statistics given in this table are given for N = 213 schools. Without the missing data replacement.

APPENDIX B

B.1 Slope Variability

Slopes of randomly selected 15% of total groups are presented:

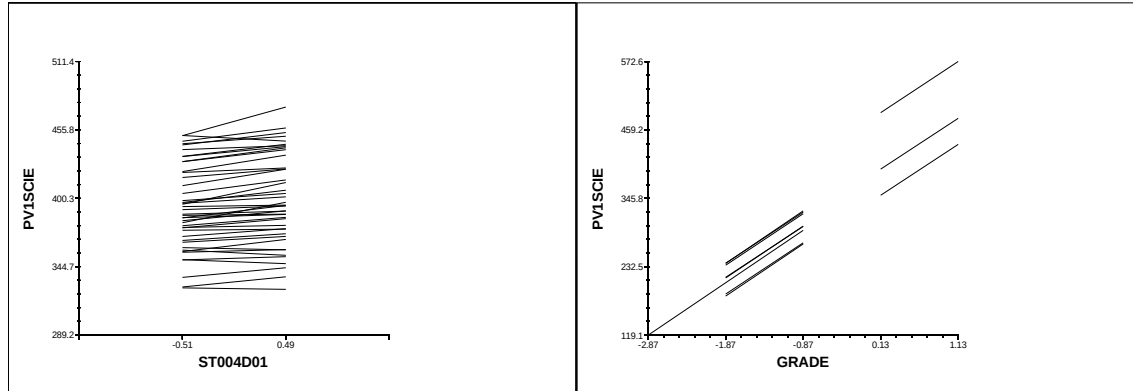


Figure B.1.1 Slope variability of ST004D01

Figure B.1.2 Slope variability of GRADE

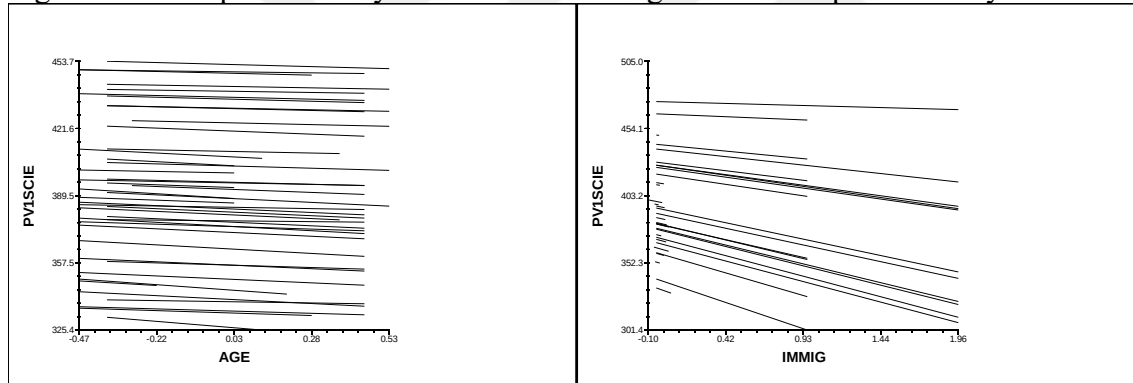


Figure B.1.3 Slope variability of AGE

Figure B.1.4 Slope variability of IMMIG

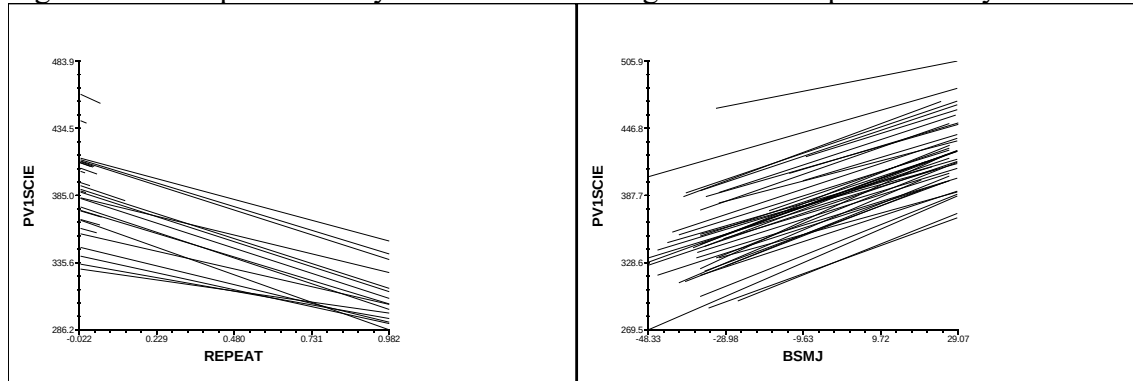


Figure B.1.5 Slope variability of REPEAT

Figure B. 1.6 Slope variability of BSMJ

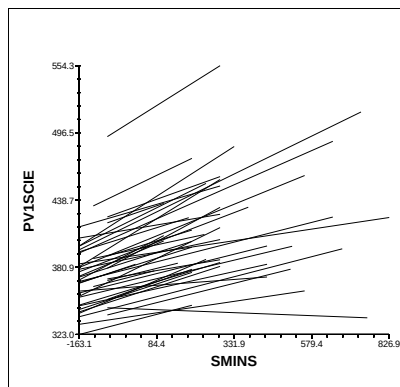


Figure B.1.7 Slope variability of SMINS

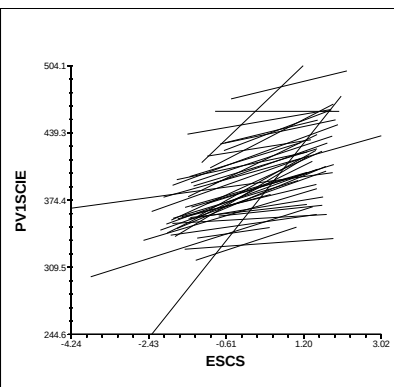


Figure B.1.8 Slope variability of ESCS

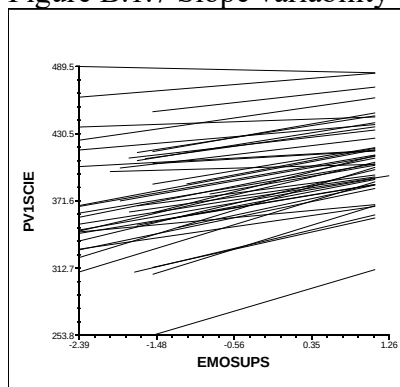


Figure B.1.9 Slope variability of EMOSUPS

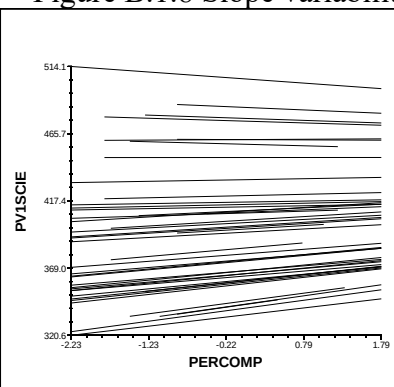


Figure B.1.10 Slope variability of PERCOMP

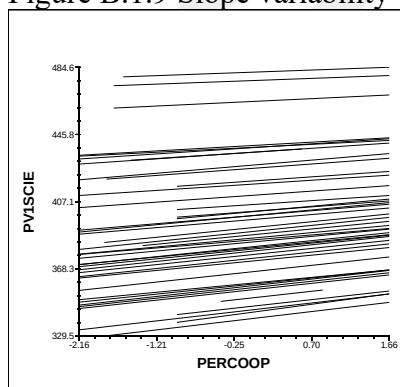


Figure B.1.11 Slope variability of PERCOOP

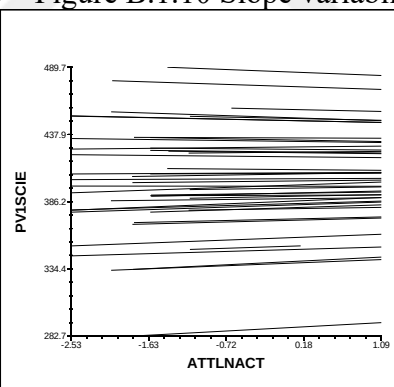


Figure B.1.12 Slope variability of ATTLNACT

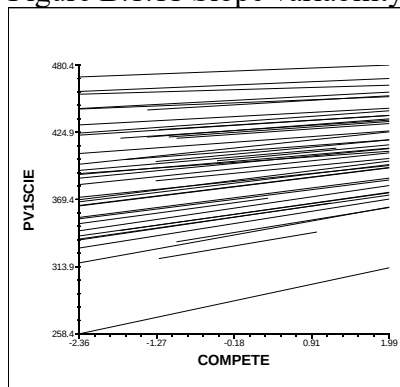


Figure B.1.13 Slope variability of COMPETE

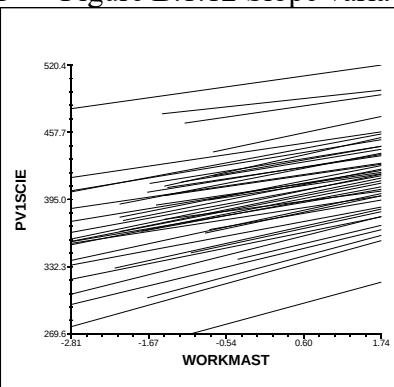


Figure B.1.14 Slope variability of WORKMAST

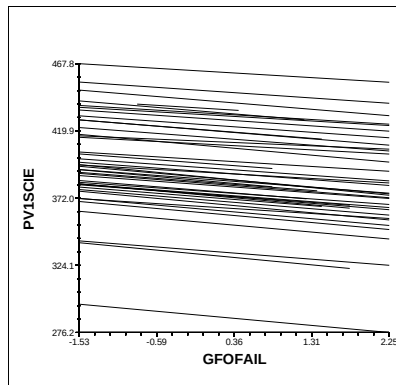


Figure B.1.15 Slope variability of GFOFAIL

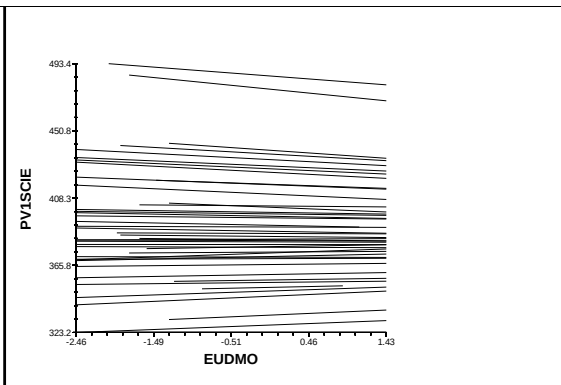


Figure B.1.16 Slope variability of EUDMO

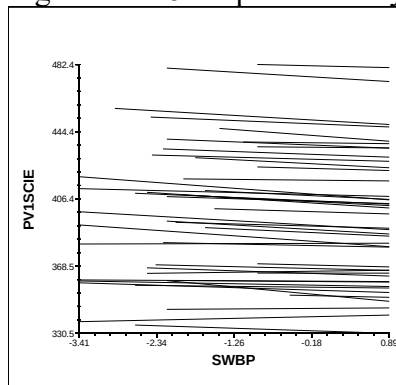


Figure B.1.17 Slope variability of SWBP

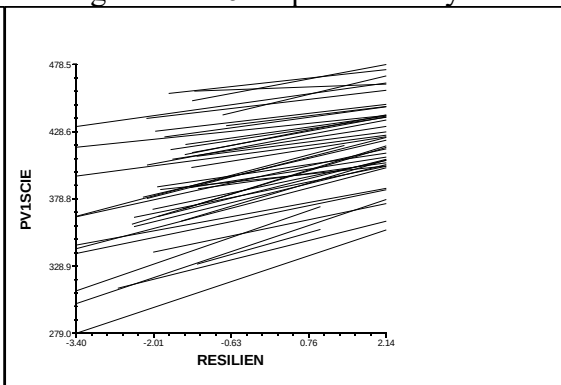


Figure B.1.18 Slope variability of RESILIEN

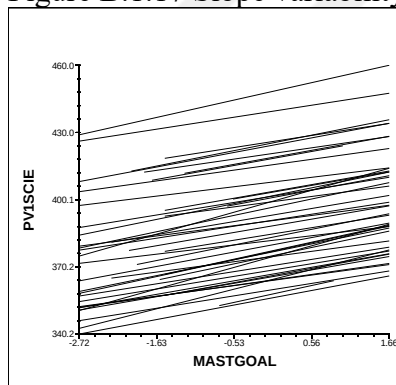


Figure B.1.19 Slope variability of MASTGOAL

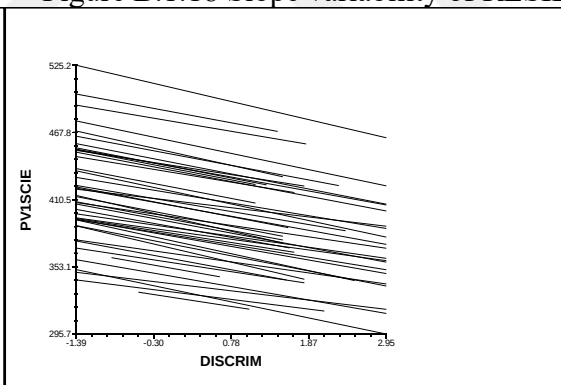


Figure B.1.20 Slope variability of DISCRIM

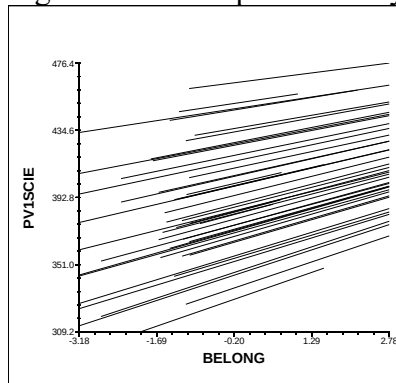


Figure B.1.21 Slope variability of BELONG

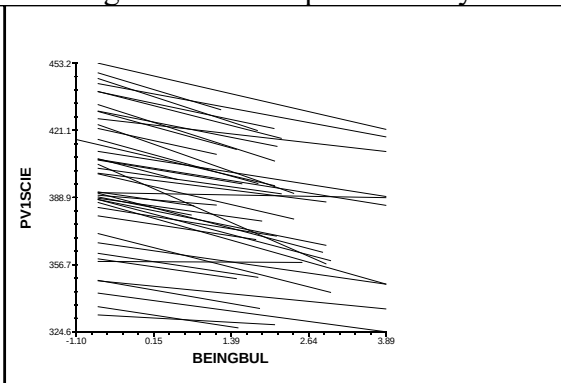


Figure B.1.22 Slope variability of BEINGBUL



Page left intentionally blank.