Olav M. Troviks vei 46
0864 Oslo
tctan@student.uv.uio.no

Friday 28 May 2021

Centre for Educational Measurement
University of Oslo
PO Box 1161 Blindern
0373 Oslo

Dear Tara and PhD Selection Committee,

Application for PhD Research Fellow in Educational Measurement

I wish to present myself for your consideration for CEMO 2021's PhD research fellow position examining the cross-national trends in educational equality and equity. The project's mission in social equality attracts my interest and passion in particular and I believe my experience and training in quantitative analyses would lend itself well to the methodological demand of this research.

Both my prior training in economics and CEMO's master program have prepared me for the pathway towards quantitative research. I received A grades for both the item response theory (sample writing in Appendix C) and multilevel modelling courses (Appendix D), two key analysis skills applicable to the PhD research project. Under the guidance of Prof Scherer and Dr Chen, I further studied and applied more recent advancement in the methodology literature to my master thesis (Appendix B) working with large-scale international assessment data. My thesis also examined gender differences and cross-country comparisons, two main research aims proposed by the PhD project.

I have been continuously active with UV's research projects in addition to my master degree candidacy. I had the privilege of working under experienced scholars across numerous groups and developed research project management skills such as literature screening, bibliography management, corresponding with authors for supplemental material, as well as coding publications into literature review databases. My own master thesis became the immediate beneficiary of my research assistant experience and I believe I can readily apply and further perfect these research skills in the PhD project.

I believe my pedagogy training and teaching experience would also serve a PhD candidate's duty well. I grew up with CEMO's master program since its inception in 2018 and was honoured to further contribute to this degree through both the Admission Committee and Student Board in 2020. I believe this unique experience would prepare me well should I be given the privilege of guiding new CEMO master students to their success.

I actively initiated and expanded professional network during my master degree candidacy. I thank EngageLab for inviting me to their Student Innovation project in late 2020, leading to the subsequent scholarship and prototype development. Terje our IT specialist also involved me in the trail and set-up of high performance computing infrastructure recently. At extra-curriculum activities, I was elected as the treasurer for the Oslostudentenes Idrettsklubb Svømming, charged with finance and administration responsibilities.

Research success also requires maintaining connections at a personal level. During my master candidacy, I have developed healthy and productive rapport with CEMO and neighbouring IPED staff while studying in the student room. I trust my team player spirit will bring pleasant interactions to the office floor should I be invited to continue my successful learning story into a PhD.

Should you desire any additional information, please do not hesitate to contact me by phone 41 22 39 75 or by email tctan@student.uv.uio.no.

Thank you for your kind consideration and I look forward to hearing from you.

Sincerely,

Tony C. A. Tan

Appendix A

Curriculum Vitæ

Tony C. A. Tan – Curriculum Vitæ

Address	Olav M. Troviks vei 46, 0864 Oslo	**Mobile Phone**	+47 - 41 22 39 75
Date of Birth	26 July 1985	**Email**	tctan@student.uv.uio.no
Nationality	Australian	**Languages**	English, Chinese (native), French (B1), Norwegian (A1)

Professional Affiliation

2019-present Registered teacher (Utdanningsdirektoratet referanse 2019/5623)
Registered teacher (Teachers Registration Board of South Australia, reg no. 615535)

Award

2020 UV Student Innovation Project

Designed an automating system for marking and performance analyses of non-standardised assessment tasks

Education

2019-2021 Master of Science (Assessment, Measurement and Evaluation) - University of Oslo

Thesis topic: Identifying school climate variables associated with financial literacy outcomes in PISA 2018 data: A multilevel structural equation modelling approach
Supervisors: Prof Ronny Scherer and Dr Chia-Wen Chen

2017-2019 Master of Teaching (Secondary) - University of Melbourne

First Class - 80% Average
Licensed teaching areas:
Victoria Certificate of Education (VCE) accounting, economics, and legal studies;
Year 7–10 economics and business, civics and citizenship, history, and geography
Exchange semester, Centre for Educational Measurement, University of Oslo
Melbourne Global Scholars Award recipient
Working with Children Check (VIC) holder
Mental Health First Aid certificate holder

2012 Taxation Law & Practice (Non-award) - University of Canberra

First Class - 95%

2008-2009 Master of Professional Accounting - Australian National University

First Class - 80% Average
Treasurer, ANU Postgraduate and Research Students' Association
Student Member, Charted Accountants Australia & New Zealand

2005-2008 Bachelor of Economics - Australian National University

2003-2004 South Australian Certificate of Education - Eynesbury College

Employment and Experience

Feb 2020 - present CEMO & ILS, UV, UiO
Research Assistant

- Student representative, CEMO Board
- Member, CEMO Admission Committee
- Participate in literature screening, bibliography management, coding and analyses in the following projects:
  - School climate review, principal investigators Prof Trude Nilsen & Dr Nani Teig
  - Science teaching and learning systematic review, principal author Dr Nani Teig
  - Executive function and the emergence of pre-schoolers' mathematics capabilities meta-analysis, principal author Mr Valentin Emslander (University of Luxembourg) and Prof Ronny Scherer

Dec 2020 - present Oslostudentenes Idrettsklubb Svømming
Økonomiansvarlig

- Manage club bank account
- Approve invoices, reimbursement and expense requests
- Maintain financial record and membership register
- Reconcile monthly accounting report
- Organise training sessions, marketing and administrative duties

May 2015 - Feb 2017 Lyo Education Pty Ltd
Chief Finance Officer

Managed company finance, taxation, and legal compliance

Sept 2013 - Dec 2014 Zhengzhou University, Zhengzhou High School (No. 18), Huanghe Science & Technology College
Lecturer and teacher

Convened the following courses:

- Financial accounting, microeconomics, academic English
- VCE accounting, economics and mathematics

Jan 2011 - Dec 2011 Australian Defence Force Academy (ADFA), School of Business
Tutor

Tutored Australian Defence Force military cadets economics, accounting and business statistics

July 2008 - Sept 2013 Accountant and business advisor
Graduate internship and private practice

References

Name	Prof Ronny Scherer	**Name**	Dr Chia-Wen Chen
Institution	CEMO, University of Oslo	**Company**	CEMO, University of Oslo
Position	Professor	**Position**	Post-doctoral Researcher
Email	ronny.scherer@cemo.uio.no	**Email**	c.w.chen@cemo.uio.no

Appendix B

Master Thesis

Identifying School Climate Variables Associated with Financial Literacy Outcomes in PISA 2018 Data

A Multilevel Structural Equation Modelling Approach

Tony C. A. Tan

Master of Science (Assessment, Measurement and Evaluation)
30 credits

Centre for Educational Measurement
Faculty of Educational Sciences

Spring 2021

Identifying School Climate Variables Associated with Financial Literacy Outcomes in PISA 2018 Data

A Multilevel Structural Equation Modelling Approach

Tony C. A. Tan

Identifying School Climate Variables Associated with Financial Literacy Outcomes in PISA 2018 Data

http://www.duo.uio.no/

Printed: Reprosentralen, University of Oslo

敬致父母

To my parents

Study hard what interests you the most in the most undisciplined, irreverent and original manner possible.

Richard P. Feynman

Contents

List of Tables

List of Figures

Acknowledgement

I am deeply indebted to my supervisors Prof Ronny Scherer and Dr Chia-Wen Chen. Both scholars afforded me their valuable time and infinite patience throughout the thesis production process. Ronny's occasional "wow" and "awesome" time and again saved me from deep insecurity and lent me the courage to continue this journey. It were Chia-Wen's clear and decisive instructions that helped me avoid the many pitfalls along the way. This thesis owes its very existence to both experts' input and reviews while any errors and omissions remain mine solely.

I often imagine what I would be doing today should former CEMO senior advisor Anne-Catherine Lehre not have encouraged me to extend my exchange semester into this Master program. I owe this degree to Anne-Catherine, Senior Executive Officer Siri Heslien, Academic Head of Program Prof Dr Johan Braeken and CEMO Director Prof Dr Sigrid Blömeke for their encouragement and guidance. My deep appreciation also goes to our new Administrative Leader Tara Sarin for providing me with a study space and her continuous support and protection throughout the pandemic period. May this thesis serve as a thank-you to everybody that assisted me each step of the way.

ILS scholars Dr Nani Teig and Prof Trude Nilsen's generosity of accepting me as their research assistant also introduced the fine details of conducting scientific research to my workflow. In fact, I learnt the school climate framework from Nani and Trude's project and I hope I presented this literature correctly in my writing.

Behind-the-scenes heroes are UV's IT expert Terje Thoresen and all the specialists empowering the Norwegian Research and Education Cloud. I can hardly imagine what my statistical analyses would look like without them lending me the university's powerful computation infrastructure. Terje also regularly stopped by and enquired about my progress—his insight into economics, politics, and life in general helped to contextualise my thesis topic into a broader social debate. I am also grateful for Terje's assistance in translating the abstract into Norwegian.

Certainly, my story with UiO would not come to be, had the University of Melbourne not facilitated my exchange semester in 2018. I would like to thank Global Learning Senior

Education Abroad Adviser Mr Aaron DeBono for opening this world to me. I look forward to reporting back to the MGSE team when international travel resumes. I wish to sit down with Aranka Dalgleish and Jeff Kinsman and highlight their thoughts and teachings that made into my thesis. The kind images of the Director of Initial Teacher Education Dr Daniela Acquaro, Student Experience Team Leader Clair Richards and Student Wellbeing Officer Ms Susie Reay, all remind me the many shoulders I stand on today and I wish to record my gratitude to you all into this permanent archive at the University of Oslo.

To mum and dad, you deserve the most special thank-you on its own page and I present this thesis to fulfil my promise to be good and study hard made at the airport 18 years ago to this day. I hope I made you proud.

我对父母的感谢自成一页。18 年前的今天，我在机场临别时曾许下诺言：我在国外会好好表现、努力学习。希望这篇论文能让二老欣慰。

Popular Abstract

Preparing youth for life-long success with numeracy, literacy and science capability has been the core mission for all schooling systems. The post-financial crises and post-COVID era, in addition, imposed increasing demand for financial literacy on school leavers. Although financial education programs were generally reported as effective in promoting learners' financial literacy outcome, paradoxical results of non-findings or even negative findings were not unheard of. Any claim that education efforts did not matter, or even harmful, for learners' development deserves immediate attention because if school were committing something wrong, school leaders and policy makers would want to know what, which and where the problems were so that harmful practices can be reverted into good pedagogy. Alternatively, it could instead be the instrument some researchers employed that led to such underwhelming results. A closer examination of how school effectiveness is measured would also promote methodology practices and the resultant policy advice. Using 2018 PISA financial literacy data, this study examined how students' financial literacy scores changed systematically as educational efforts, parental involvement, school safety as well as resource allocation varied. Analyses showed that all four aspects of school climate mattered greatly in explaining differences in students' financial literacy scores. Negative results reported by some papers were likely the results of certain design choices. School financial education should definitely not be withdrawn but more carefully designed with increase emphases on students' financial problem-solving skills in addition to knowledge and confidence training.

Abstract

Repeated financial crises and the current pandemic emergency all exposed the harsh consequences of financial illiteracy shared by large proportions of the general population. Although remedial plans were shown to be most effective if introduced early in life, the exact relationships among student-, family- and school-factors behind youth's financial literacy outcomes were not yet fully understood. Using the latest Programme for International Student Assessment (PISA) 2018 financial literacy data and the theoretical framework of school climate recently proposed by Wang and Degol (2016), this study examined the mechanism for individuals' financial literacy performance in the context of their school environment. A multilevel structural equation model (MSEM) revealed that 33.5% of the variation in students' financial literacy scores could be explained by student-level variables and 47.7% by school-level factors for the full PISA 2018 sample. The MSEM also highlighted key roles financial knowledge and financial confidence played in mediating students' financial literacy performance. Both financial education and financial socialisation were positively associated with financial knowledge and confidence, but their direct effects on financial literacy scores were negative once the mediation effects have been accounted for. Strong contextual effects suggested the important role of school environment for facilitating individual-level effects. This study took a person-ecological approach for reconciling two strands of research efforts that focused either on students or on schools. It also confirmed the importance of school education, parental involvement, safety and educational resources for bringing about greater financial knowledge and confidence and identified potential improvement opportunity for pedagogical practices for further advancing students' financial problem-solving capabilities.

Keywords: school climate, financial literacy, PISA, multilevel modelling, structural equation modelling, contextual effect

Journal of Economic Literature Classification: A21, C13, C31, I21

Abstrakt

Gjentatte finanskriser og den nåværende pandemisituasjonen avslørte de alvorlige konsekvensene av manglende økonomisk kunnskap i en betydelig andel av befolkningen. Selv om kompenserende tiltak har vist seg å være mest effektive ved introduksjon tidlig i livet, var de eksakte forholdene mellom student-, familie- og skolefaktorer vedrørende ungdoms økonomiske ferdigheter ikke helt forstått.

Ved hjelp av det nyeste programmet for internasjonal studentvurdering (PISA) 2018—økonomiske ferdigheter og det teoretiske rammeverket for skoleklima, som nylig ble publisert av Wang og Degol (2016), undersøkte denne studien mekanismen for individers økonomiske ferdigheter i skolesammenheng. Strukturell flernivåmodellering (MSEM) avslørte at 33,5% av variasjonen i studentenes økonomiske ferdigheter kunne forklares med variabler på studentnivå og 47,7% av faktorer på skolenivå for hele PISA-utvalget. MSEM fremhevet også nøkkelroller som finansiell kunnskap og økonomisk tillit sin betydning i formidling av elevenes økonomiske ferdigheter.

Både finansiell utdannelse og økonomisk sosialisering var positivt assosiert med økonomisk kunnskap og tillit, men deres direkte effekter på finansiell kompetanse var negative etter at meklingseffektene har blitt redegjort for. Sterke kontekstuelle effekter belyste skolemiljøets viktige rolle for å tilrettelegge effekter på individnivå. Denne studien tok en personøkologisk tilnærming med formål om å forene to forskningsfelt som fokuserte på enten studenter eller skoler. Den bekreftet også viktigheten av skoleundervisning, foreldrenes engasjement, sikkerhet og pedagogiske ressurser for å skape større økonomisk kunnskap og tillit, og identifiserte potensielle forbedringsmuligheter for pedagogisk praksis for å videreutvikle elevenes økonomiske problemløsningsferdigheter.

Nøkkelord: skoleklima, finansiell forståelse, PISA, flernivåmodell, strukturell ligningsmodell, kontekstuell effekt

JEL-klassifisering: A21, C13, C31, I21

论文摘要

多次的金融危机和当前的新冠大流行都暴露了大量的普通民众由于财经素养缺乏而导致的严峻后果。尽管有先前文献报道补偿财经素养应尽早介入，但学术界尚不完全清楚学生、家庭和学校诸因素是如何联合影响孩子的财经素养培育的。本论文汲取最新的 2018 国际学生评估项目（PISA）财经素养数据，结合 Wang 和 Degol（2016）共同提出的学校氛围理论框架，对学生是如何在学校大环境下养成财经素养这一机制进行了研究。多层结构方程模型（MSEM）显示，在 2018 年 PISA 样本中，33.5%的财经素养得分差异可以由学生层面的变量解释，另外 47.7%可以由学校层面的因素解释。MSEM 还显示财经知识和财经信心对学生的财经素养得分表现起了中介作用。财经教育和财经浸润都与财经知识和信心成正相关。但是一旦控制了中介效应，教育和浸润对财经素养得分的直接影响则降为负值。显著的情境效应表明学校在学生素养养成过程中起了重要作用。现有文献要么以学生为研究对象，要么以学校为分析单元来理解财经素养培养。本文则采用了个人-生态观点对二者进行了调和。本论文还确认了学校教育、父母参与、校园安全和教育资源对于提高财经知识和信心的重要性，并对教学实践提出了一些改进建议以进一步提高学生的财经问题解决能力。

关键词：学校氛围；财经素养；国际学生评估项目（PISA）；多层模型；结构方程模型；情境效应

《经济文献杂志》分类：A21, C13, C31, I21

Chapter 1 Introduction

1.1 An Atlas of Financial Illiteracy

Repeated economic crises in recent memory have exposed the harsh consequences of financial *illiteracy* shared by high proportions of the general population. Low financial literacy was directly linked with negative credit behaviours such as high amount of credit card debt (Norvilitis & MacLean, 2010), high costs of borrowing (Huston, 2012; Pak, 2018), poor mortgage choices (Cox et al., 2015) and subsequent delinquency and home foreclosure (Agarwal, Chomsisengphet et al., 2015; Gerardi et al., 2010). Poor financial decisions made early in life can have profound long-term economic and societal impacts (Montoya & Scott, 2013) such as forgoing medical care (Lusardi et al., 2015), mental health crises (Stone et al., 2018) and geronto-poverty resultant from insufficient retirement provision (Lusardi & Mitchell, 2007, 2008). Borrowers' collective misjudgement on mortgage risks kicked start the subprime crises and in combination with Wall Street greed and laissez faire regulatory attitudes that eventually triggered the avalanche of 2008 financial crisis, the first domino of world-changing events whose impact continues reshaping global economics and geopolitics landscape.

Even more concerning is the pervasive global distribution of financial illiteracy. Deficiencies in financial capability had been observed not only in emerging economies (Karakurum-Ozdemir et al., 2019) such as Colombia (Cao-Alvira et al., 2020), Mexico (Arceo-Gómez & Villagómez, 2017; Böhm et al., 2021), India (Agarwal, Amromin et al., 2015; Kiliyanni & Sivaraman, 2016; Utkarsh et al., 2020), Indonesia (Cole et al., 2009; Khoirunnisaa & Johan, 2020), Turkey (Akben-Selcuk & Altiok-Yilmaz, 2014), and Eastern European countries (Belás et al., 2016; Opletalová, 2015; Reiter & Beckmann, 2020) but also in advanced economies such as Australia (Ali et al., 2014; Taylor & Wagland, 2013; Thomson & De Bortoli, 2017), Canada (Boisclair et al., 2017), Germany (Bucher-Koenen et al., 2017; Erner et al., 2016), Austria (Silgoner et al., 2015), the UK (Barnard et al., 2021) and the USA (Breitbach & Walstad, 2016; Gale et al., 2012; Lusardi et al., 2010). International comparisons also reported low financial literacy in many Asian countries (Yoshino et al., 2015) and member states of the Organisation for Economic Co-operation and Development (OECD) (Cupak et al., 2018; Lusardi, 2015), particularly amongst the young (De Beckker et al., 2019), females, lower educated (Klapper & Lusardi, 2019) and somewhat surprising, inhabitants of countries with more generous social security systems (Jappelli, 2010).

1.2 Financial Literacy as a Necessity

One major reason behind the escalating interests in citizens' financial literacy can be attributed to the policy adjustment taking place in the past two decades. The neo-liberal ideology of reducing government involvement in the economy had crowded out societal care such as pension, health and education from the collective via the state to the individuals (Gilbert, 2002). In a post-financialisation world (Krippner, 2005), the primary goal of political economy has shifted from the redistribution of wealth to the incorporation of individuals within the mainstream financial architecture (Regan & Paxton, 2003). The succession of the asset-based welfare system to the income-based model (Finlayson, 2009), however, was by no means unique to the Anglosphere. The Hartz reforms of 2003/04, according to Seeleib-Kaiser (2016), had significantly altered Germany's post-war social welfare arrangement, leading Ferragina et al. (2015) to re-classify Germany from a conservative welfare into a liberal welfare state comparable to the United Kingdom. Although a detailed account of the history, politics and moral philosophy of social welfare reforms is beyond the scope of this project, this background information does confirm financial literacy as a social necessity independent of one's believes or preference.

Strengthening citizen's financial literacy also generates substantial social returns. The latest U.S. Department of Justice statistics showed a total loss of near 3.25 billion dollars to financial fraud in 2017 (Morgan, 2021) while similar figure was estimated to be 190 billion pounds for the UK, more than the public spending on health and defence *combined* (Gee, 2018). A financially informed and alert individual is less likely to fall victim to fraud and scams (Gamble et al., 2015; Lusardi, 2012) although this effect was thought to be moderated by one's ability to recognise and resist manipulative tactics (Drew & Cross, 2016). In addition to the monetary benefit, some scholars see financial education as a service to civics and democracy since a financially literate population is more resilient to political opportunists. Teaching citizens—as well as the young who will be future voters—about taxation, tariff, outsourcing, labour market transition and career choices protects not only individuals' financial security and dignity but also informs and empowers voting behaviours through which governments are scrutinised and democracy is upheld (Davies, 2015) and even modified (Arthur, 2016). After all, financial literacy can be seen as an investment in human capital (Lusardi & Mitchell, 2014). Today's young people are growing up in a society in which the financial landscape is complex and the financial responsibilities of citizens are substantial.

1.3 Profiles of Successful Learners

As the cellular constituent of the broad economy, personal finance success has long attracted interests from policy makers and educators. Numerous research efforts have been devoted into identifying the common traits shared by individuals displaying knowledge, confidence and behaviour conducive to high financial literacy performance. Potrich, Vieira and Kirch (2015) found well-educated individuals from wealthy families and earning good income themselves had the highest propensity to demonstrate substantial financial literacy. The positive correlations between socioeconomic status and financial literacy performance was observed not only in adult samples but also in late year school students. Using school enrolment data from the State of Victoria, Australia, Ali et al. (2016) found socio-economic variables such as urban-rural locations, non-English speaking at home as well as parental education and occupations accounted for very high proportion of the variations in students' financial literacy test scores. Negative correlations, on the other hand, had been observed between cross-border relocation experience and financial literacy performance. Using 2012 PISA data, Gramaţki (2017) applied a propensity score matching technique to 15-year-old migrant students and concluded that, everything else being equal, second generation migrants underperformed their native peers by 0.15 standard deviations (SD) and this penalty increased to 0.30 SD for first generation migrants.

In addition to social factors, there appeared to be a persistent and sizeable sex difference in financial literacy performance with greater awareness of monetary matters amongst males (Atkinson & Messy, 2011; Lusardi et al., 2010) regardless of test question sophistication (Agnew & Cameron-Agnew, 2015; Agnew & Harrison, 2015) and across countries (Bucher-Koenen et al., 2017). Correlational studies largely discounted macroeconomic variables behind male advantages in financial literacy performance (Chambers & Asarta, 2018) in favour of factors at the family level (Chambers et al., 2019), corroborating the observation that females appeared to start falling behind too early in life (Driva et al., 2016) to allow market force to take effect (Preston & Wright, 2019). Culture did seem to play a partial role in explaining sex difference (Grohmann, 2016) with gender gaps appearing significantly smaller in countries with more egalitarian financial arrangement for custody and marriage (Hospido et al., 2021). Additional proposals were also put forward ranging from historic forces (Bottazzi & Lusardi, 2020), risk aversion (Chen & Garand, 2018), lacks of confidence (Bucher-Koenen et al., 2021; Danes & Haberman, 2007) or problem-solving attitudes (Longobardi et al., 2018), to imbalanced household decision-making (Fonseca et al., 2012). Consensus remains strong amongst existing literature advocating more inclusion of women in promoting population's financial literacy and well-being.

1.4 Measuring Financial Literacy

All intervention programs aiming for financial literacy advancement must be constructed based on sound evidence. Amongst competing inventories, OECD's Programme for International Student Assessment (PISA) stands out as a comprehensive and reliable source of data for measuring 15-year-olds' financial literacy outcomes thanks to OECD's careful sampling procedure and attention to construct validity of measurement. Four technical features of PISA are crucial for the architecture of this study. First, following statistical theory, PISA designers acknowledged the hierarchical nature of education research data such that students are nested in schools, and schools are further nested in countries. Second, one student weight is assigned to each observation in order to account for the fact that not all schools in a country are equally likely to be sampled by the PISA organiser; and given a particular school that has been chosen, not every student in this school is equally likely to be asked to participate in the test (Rust, 2014). A third complication arises from the "planned missingness" in students' responses because each participant is only given a small number of questions relative to the entire test bank in order to ensure their responses are not undermined by tiredness (von Davier, 2014), leading to the outcome variables being represented by multiple plausible values. Fourthly, PISA consulted and synthesised multiple schools of thoughts (OECD, 2019a) in constructing their financial literacy framework. As a result, 2018 PISA data set (OECD, 2020a) provides not only variables measuring behavioural competency outcomes but also cognitive and affective factors such as familiarity with concepts of finance and confidence about financial matters, enabling a nuanced study design involving decomposing the total effect of financial literacy performance into its knowledge, affect, and application components.

1.5 Program Effectiveness for Advancing Financial Literacy

Since youths partition their time between schools and families, research efforts aimed at promoting young people's financial literacy over the years evolved into two strands: on the design and evaluation of school financial education programs, and on the influence of home environment through the process of financial socialisation—the intentional or involuntary transmission of financial concepts which are required to functioning successfully in society (Bowen, 2002). A recent meta-analysis conducted by Kaiser and Menkhoff (2020) found that while school financial education programs had sizeable impacts on *financial knowledge* (+0.33 *SD*) similar to education interventions in other domains, their effect on students' *financial behaviour* is quite small (+0.07 *SD*). This conclusion added to a list of weak or non-findings regarding the long-term behavioural effect brought about by school financial education programs. Brown et al. (2016), for instance, reported mixed outcome in students' long-term financial well-being depending on the programs received; whereas Cole et al. (2016) observed that traditional personal finance courses lacked any explanatory power in accounting for graduates' financial outcome once the additional mathematics training in which finance topics were packaged has been controlled for. Despite careful controls and thoughtful study designs, correlating classroom interventions and young people's financial literacy outcomes has repeatedly yielded paradoxical results of non-significant or even negative relationship; some positive findings remained small in magnitudes and/or were sensitive to robust analyses.

Literature along the financial socialisation line of enquiry delivered more consistent findings. Building on the acknowledgement that families serve as information filters from the outside world (Danes & Haberman, 2007) as well as the foundation for youth's continued financial concept formation, Gudmunson and Danes (2011) put forward a family financial socialisation theory to accommodate both the process and the outcome for variations in young people's financial capabilities. Using structural equation modelling, Jorgensen and Savla (2010) was able to show that perceived parental influence had a direct and moderately significant influence on financial attitude, did *not* have an effect on *financial knowledge*, and had an indirect and moderately significant influence on financial behaviour, mediated through financial attitude. This attitude(A)–behaviour(B)–cognition(C) conceptualisation of financial literacy (Potrich, Vieira, Coronel et al., 2015) continues to influence subsequent research effort. More recently, Moreno-Herrero et al. (2018) continued this line of enquiry by applying multilevel regression analyses to the 2015 PISA data and reported that students' financial literacy was associated mainly with understanding the value of saving and discussing money matters with parents. In addition, exposure and use of financial products, in particular holding a bank account, improved students' financial knowledge as well.

1.6 Research Questions

The current study wishes to incorporate both the school intervention and family socialisation arms of existing literature under a uniform framework recently proposed by Wang and Degol (2016) named "school climate". Besides the classroom activities (ACADEMIC) and parental involvement (COMMUNITY) aspects reviewed earlier, the school climate framework also acknowledges the importance of school safety (SAFETY) and adequate resources (INSTITUTIONAL ENVIRONMENT) for cultivating a healthy and thriving young generation. By taking advantage of the latest wave of 2018 PISA financial literacy results, this project aims to answer these two research questions:

RQ1. To what extent can the variation in students' financial literacy outcomes be accounted for by each of the school climate variables?

RQ2. How does the school-level climate impact on individual learners' financial literacy acquisition process?

1.7 Thesis Overview

This thesis is structured as following: Key concepts such as school climate and financial literacy are explained in detail in Chapter 2 along with the hypothesised relationship between each construct. Chapter 3 will explain the 2018 PISA financial literacy data including sample characteristics and variable formation. A multilevel structural equation model will be proposed in this chapter as well as related technical considerations such as weights, estimators and the model evaluation procedure. Subsequently, analysis results will be presented in Chapter 4 including both descriptive and inferencial statistics. Coefficients from student- and school-levels will be presented separately first, then linked together by the contextual effects. Finally, Chapter 5 will discuss the pedagogical and policy implications of these findings, pointing out the limitation on causal inference as well as directions for future research effort.

Chapter 2 Conceptual Framework

2.1 School Climate

A positive school climate is easier to recognise but difficult to define (OECD, 2019b). When organising school attributes into frameworks, early studies loosely clustered themselves into two camps along the concrete–abstract spectrum. When researching on students' behavioural problems and emotional distress, for example, Kuperminc et al. (1997) recognised the insufficiency of using observable characteristics of a school as the metric for its managerial success but adopted a utilisation and perception approach based on social-ecological and developmental theories. Such emphasis on school users' *perception* continued into Esposito (1999)'s study of students' social disadvantages on their academic outcomes, with exploratory factor analysis results suggesting a five-factor model including student academic orientation, parent-school relationships, security, administration and teacher-student relationships. Freiberg and Stein (1999), on the other hand, took a more idealised view of school climate as "the heart and soul of a school"—the very "essence of a school that leads a child, a teacher, an administrator, a staff member to love the school and to look forward to being there each school day" (p. 11). However broad or narrow the definition, both ends of the spectrum signalled that the ultimate utility of any school climate framework should facilitate our understanding of student development.

With this goal in mind, Wang and Degol (2016) surveyed six theories for the purpose of building a multidimensional school climate framework. Since schooling is an interaction between individuals and every environment immersing them (the bio-ecological theory), students inevitably develop protective and/or maladaptive behaviours (risk and resilience perspective) in addition to all existing bonds they formed with parents (attachment theory). Thanks to students' ever-growing capabilities, schools may then encourage learners to connect, invest, participate and believe in their learning environment (social control theory), by bridging their motivation towards success criteria (social cognitive theory) and by removing barriers (stage-environmental fit theory) to growth. These theories jointly guided a literature review and coding exercise that led to a four-domain, 13-dimension structure of school climate framework (see Figure 1, Wang & Degol, 2016, p. 318). This current project approached Wang and Degol's

(2016) ontology from the domain-level and referred the ACADEMIC climate as the overall quantity and quality of the teaching-learning activities; COMMUNITY as the engagement and interpersonal ties schools maintain with stakeholders such as and in particular parents; SAFETY as the degree of physical and emotional security afforded by schools; and INSTITUTIONAL ENVIRONMENT as the organisational and structural features of schools in particular their educational resource availability. All four branches of the school climate framework serve as platforms upon which students' financial literacy can be constructed.

2.1.1 School Financial Education Programs (FEdu)

Amongst the many redress schemes aimed at promoting citizens' financial capability, the return on investment was the highest when direct classroom interventions were applied to the young. Lusardi and Mitchell (2014) have shown that providing financial knowledge to high schoolers before they enter the labour market increased their well-being by approximately 82% of their initial wealth, while the rate of return was around 56% for college graduates. In order to test the causal effects between classroom interventions and students' financial understanding Amagir et al. (2018) reviewed 24 studies evaluating the effectiveness of secondary school financial education programs using either random control trails or quasi-experimental research designs, and found all but two reported positive effects between school interventions and students' financial knowledge. The effect sizes, however, appeared to be dependent on the length of the delivery periods, with one long and intensive program yielding $d = 0.981$ for basic economic knowledge and 1.020 for personal finance but only $d = 0.221$ to 0.267 from a short series. The review paper also found general positive correlations between school programs and students' attitudes towards finance-related matters (FA) such as confidence. Kaiser and Menkhoff (2020) recently updated the literature using publications employing (quasi-)experiment designs and reported an average treatment effect of 0.331 for the 31 pooled samples and 0.369 for the 12 high school sub-samples on financial knowledge (FC) gains. Based on existing literature, the current project therefore hypothesises that

H1: There exists a positive association between FEdu and FC.

H2: There exists a positive association between FEdu and FA.

The relationships between school financial education programs and students' subsequent financial *behaviours* (FB), on the other hand, were more mixed. Early studies by Bernheim et al. (2001) examined the impact of the progressive introduction of financial curriculum mandates in many US states between 1957 and 1985 on recipients' saving behaviour and net worth at the end of 1995. Analyses showed that (a) systematic differences in saving rates across states did not appear until after mandates were imposed, (b) saving rates only started to raise many years after the mandate, and (c) net worth was higher by roughly one-year's worth of earnings for an average individual having been exposed to the mandate. This 20-year time horizon study led the authors to the conclusion that school financial education efforts *did* have meaningful impact on recipients' life-long financial well-being albeit with significant implementation lags. Most recently, a German study showed causal evidence that teaching financial literacy to 16-year-olds had significant short- and longer-term effects on risk and time preferences (Sutter et al., 2020). This result lent weight to an earlier randomised controlled trial with 3,000 Grade 9 students in Spain (Bover et al., 2018) where students showed more patience in hypothetical saving choices both immediately after the treatment and three months later. Frugality, delayed gratification, faster debt clearance and decreased reliance on credit financing were all documented by Carlin and Robinson (2012) in the US after a finance-related theme park training. Other publications, however, showed weak or even non-findings for financial behaviour improvement. A short financial education program on German high schoolers, for example, showed reduction in impulse purchases but no significant increase in savings (Lührmann et al., 2015). A review article by Fernandes et al. (2014) found school programs explained only 0.1% of the variance in financial behaviours and decaying to negligible levels 20 months later. Since the current literature is yet to reach consensus about the strength of the relationship between school interventions and students' financial behaviour, it is prudent to hypothesise:

H3: The relationship between FEdu and FB is non-negative.

2.1.2 Parental Influence and Financial Socialisation (FSoc)

Although financial capability is an important integral of adulthood, the process of acquiring the financial knowledge and skills begins in early childhood. Parents provide a context in which children learn what money is, for instance, and how it is used and saved (Birbili & Kontopoulou, 2015). Whether intentionally or informally, financial intuition is passed around the household through frequent interactions, conversations, and lessons. Consequently, the financial knowledge and skills acquired while growing up at home form the foundation for the financial attitudes and behaviours carried into adulthood (Serido & Deenanath, 2016). Using a panel data set from the Dutch DNB Household Survey between 2000 and 2012, Bucciol and Veronesi (2014) reported that parental teaching about savings increased the likelihood of adult saving by 16% and the saving amount by approximately 30%. Similar intergenerational effect was observed from longitudinal studies in the US, linking adolescents' observation of parents' responsible financial behaviour to their own good decisions and actions later in life (Tang, 2017). Moreno-Herrero et al. (2018) further examined the relationship between students' financial socialisation experience and their financial literacy outcome using PISA 2012 data. By operationalising financial socialisation as the frequency of money-related discussions with parents, saving habits and bank account ownership, the authors reported positive associations between financial socialisation and PISA financial literacy scores. These studies suggested that

H4: The relationship between FSoc and FC is non-negative.

H5: FSoc is positively related to FA.

H6: FSoc is positively related to FB.

2.1.3 School Safety (Safety)

School safety is the prerequisite for any learning and growth. As a social construction, the definition of school safety can be subjective and coloured by one's social location, cultural experiences and school context (Cornell & Mayer, 2010). Since its initial definition as an absence of weapons and/or homicides in school settings (Skiba et al., 2006), the understanding of school safety has evolved substantially to emphasise the prevention of overt and covert violence such as bullying behaviours (physical safety, Jimerson et al., 2012), caring and supportive staff as well as the availability of mental health services (emotional safety, Kuperminc et al., 1997), and delinquent acts committed by students against their peers and teachers (school order and discipline, Gottfredson et al., 2005). Although studies specifically examining the relationship between adverse school experiences such as being bullied and financial literacy performance were yet to emerge, Kutsyuruba et al.'s (2015) review article on the associations between school safety and students' general academic attainment may serve as a general guide suggesting

H7: There is a positive association between Safety and FC.

H8: There is a positive association between Safety and FA.

H9: There is a positive association between Safety and FB.

2.1.4 Institutional environment (Resource shortage)

Both the physical and social infrastructure of schools greatly influence users' experience and functioning. An optimal learning environment requires appropriate heating and cooling, ample supply of lighting, necessary acoustic control and regular maintenance (environmental adequacy, Uline & Tschannen-Moran, 2008). Secondly, structural organisation such as class size was also linked to students' education outcomes (Finn & Achilles, 1999). Lastly, although the core of classroom instruction involves the interaction between teachers and students, the quality of such interaction is frequently facilitated by the equipment, materials, and supplies. Optimising resource utilisation has been attributed to improved student attainment particularly for schools in impoverished communities (Miles & Darling-Hammond, 1998). Based on the observed impact school resource had on learner outcomes, this study hypothesises that

H10: Resource shortage is negatively associated with students' average FB.

H11: Class size is negatively associated with students' average FB.

2.2 Financial Literacy

In its official publication *PISA 2018 Assessment and Analytical Framework* (OECD, 2019a), the OECD provided an explicit definition of "financial literacy" as

> the knowledge and understanding of financial concepts and risks, and the skills, motivation and confidence to apply such knowledge and understanding in order to make effective decisions across a range of financial contexts, to improve the financial well-being of individuals and society, and to enable participation in economic life (p. 128)

with emphases on both the thinking and behaviour that characterise such construct and the purposes for developing this particular literacy. Of particular relevance to the current project are the knowledge, confidence and application aspects of financial literacy.

2.2.1 Knowledge Aspect of Financial Literacy (FC)

Since poor financial behaviours have been associated with a lack of financial knowledge (Hastings et al., 2013; Lusardi & Mitchell, 2014), one major goal of financial literacy interventions is to ensure students receive the information and support they need to make responsible and appropriate financial decisions confidently, both in their school years and in adult lives (OECD, 2020b).

2.2.2 Confidence Aspect of Financial Literacy (FA)

The positive association between students' confidence and their academic attainment has also been well documented. By synthesising one decade of large-scale international assessment data, Lee and Stankov (2018) found self-beliefs (labelled "self-efficacy" in PISA and "confidence" in TIMSS) to be the strongest non-cognitive predictor for students' mathematics achievement. Similar relationships had also been observed in the realm of financial literacy such as Arellano et al.'s (2014) study using the Spanish portion of the PISA 2012 financial literacy data, and Borges Ramalho and Forte's (2019) results based on the Brazilian sub-sample of the 2016 OECD/INFE International Survey of Adult Financial Literacy Competencies.

2.2.3 Application Aspect of Financial Literacy (FB)

Although financial knowledge and confidence forms the very foundation upon which financial capability can be developed, it is individuals' willingness and ability to *apply* such capability through financial decision-making that counts as the ultimate outcome of their financial literacy (Huston, 2010). Operationalise financial behaviour as one's ability to solve real-world financial problems also make it feasible to capture financial behaviours within a one-hour test, with the result reflecting one's understanding, affinity and application of their financial capability. The OECD paid particular attention to upholding financial literacy as an independent construct. Such consideration was important because one's financial capability was known to covary with both numeracy (Geiger et al., 2020; Ozkale & Erdogan, 2020a, 2020b; Sole, 2014) and literacy (Bay et al., 2014) skills. Empirical studies using diverse samples from the Philippines (Indefenso & Yazon, 2020) to Sweden (Skagerlund et al., 2018) reported correlations between numeracy and financial knowledge/literacy to be between approximately .61 and .52. In order to minimise the impact of low arithmetic skills (Huston, 2010), financial formulæ were never required in any problem solving tasks and students may use the on-screen calculator at any time of the test. Furthermore, stimulus material and task statements were generally designed to be as clear, simple and brief as possible to minimise the impact of low reading ability on financial literacy scores.

Both financial knowledge and confidence are hypothesised to contribute to students' performance in finance-related problem solving:

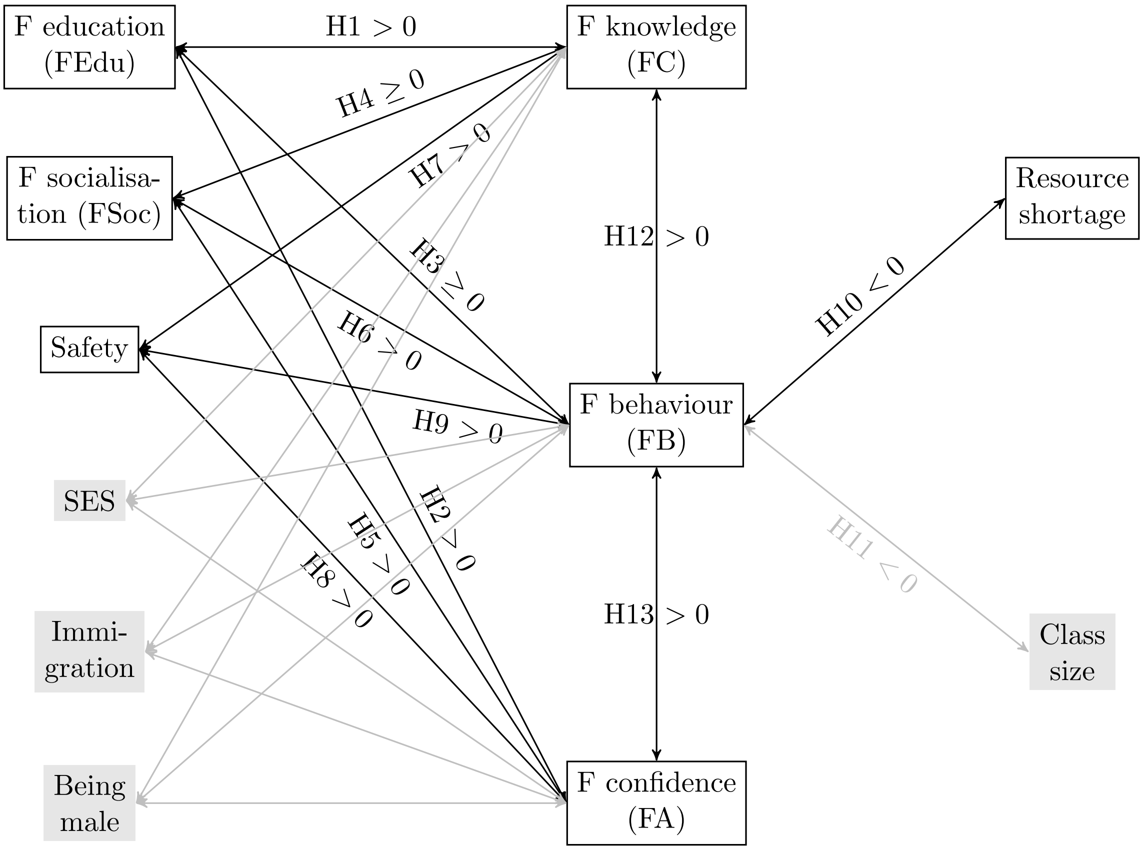
H12: FC is positively related to FB.

H13: FA is positively related to FB.

2.3 Summary of Relationships between Constructs

As discussed in Section 1.3, learners' demographic attributes such as socio-economic status, immigration history and sex were used as control variables, leading to the following diagram summarises all hypothesised relationship between concepts introduced in this chapter:

Figure 2.1
Summary of Study Hypotheses



Note. "F" is short for "Financial". Demographic control variables are shaded in grey and may covary with some or all of FC, FB, and FA.

Chapter 3 Methods

3.1 Sample

This study drew its primary data source from OECD's PISA 2018 database. Responses from both student (OECD, 2020a) and school questionnaires (OECD, 2020d) were captured and merged into a master data file using R's (Version 4.0.5, R Core Team, 2021) `intsvy` package (Version 2.5, Caro & Biecek, 2017) (see Section B.1 for analysis code) including the following 20 participating countries[1]: Brazil, Bulgaria, Canada, Chile, Estonia, Finland, Georgia, Indonesia, Italy, Latvia, Lithuania, the Netherlands, Peru, Poland, Portugal, Russian Federation[2], Serbia, Slovak Republic, Spain, and the USA. Twelve observations without school weights were dropped, leading to a sample size of 107,162 students nested in 6,631 schools (see Table B.1 for detailed sample profile). Under PISA 2018 sampling design, all student candidates were born in the year 2002 in international grades 7 or higher (Chapter 4 of *PISA 2018 Technical Report*, OECD (2020c), p. 29) and will be referred to as "15-year-old" in this study.

3.2 Measures

3.2.1 School Climate Variables

Following Wang and Degol's (2016) framework, this study selected variable `FLSCHOOL` "financial education in school lessons" as an indicator for the ACADEMIC domain of school climate; `FLFAMILY` "parental involvement in matters of financial literacy" for the COMMUNITY engagement dimension (i.e., "financial socialisation"), `NOBULLY` (reverse coding of `BEINGBULLIED` such that larger numbers imply safer schools) as an indicator for school SAFETY, and lastly `EDUSHORT` "shortage of educational material" as an indicator of the resource availability aspect of the INSTITUTIONAL ENVIRONMENT of schools. All four measures were derived variables based on IRT scaling, with good scale reliabilities for most countries and constructs (see Table B.2 for Cronbach's alphas). In addition, the OECD has applied multi-group concurrent calibrations to all latent constructs using the root mean square deviance below 0.3 criterion (for a technical discussion on RMSD, see Buchholz & Hartig, 2019, p. 244) in order to ensure cross-country measurement invariance (see Chapter 9 of *Technical Report* (OECD, 2020c, pp. 14–15) for

[1]Australia also participated in the 2018 PISA financial literacy test but chose to withhold its data from public release and is therefore not included in the current study.

[2]Moscow Region (`CNTRYID` = 982) and Tatarstan (983) have been merged into Russian Federation (643).

Table 3.1
Summary of Measures and Variables

Analysis level	Exogenous variable		Endogenous variable	
	School climate (Input, X)	Demographic control	Financial capability indicators FC & FA (Mediator, M)	FB (Outcome, Y)
School-level ($L2$)	$\texttt{FLSCHOOL}_B$ $\texttt{FLFAMILY}_B$ $\texttt{NOBULLY}_B$ `EDUSHORT`	`STRAIO`		$\texttt{FLIT}_B$

Student-level ($L1$)	$\texttt{FLSCHOOL}_W$
	$\texttt{FLFAMILY}_W$
	$\texttt{NOBULLY}_W$
		`ESCS`
		`IMMI1GEN`
		`IMMI2GEN`
		`MALE`
			`FCFMLRTY`
			`FLCONFIN`
				$\texttt{FLIT}_W$

Note. The within- and between-level components are marked with subscript $_W$ and $_B$ respectively.

analytical details).

3.2.2 Financial Literacy Measures

Financial Knowledge (FC)

In order to ascertain candidates' current understanding of finance-related topics, `FL164` of the financial literacy questionnaire presented 18 terminologies such as exchange rate, budget, and income tax and asked students to rate their familiarity with each term using a three-point scale: "Never heard of it", "Heard of it, but I don't recall the meaning" and "Learnt about it, and I know what it means". Sum scores of `FL164` were used to construct "familiarity with concepts of finance" variable (`FCFMLRTY`, Chapter 16 of *PISA 2018 Technical Report*, OECD (2020c), p. 23). This scale had good reliability properties evidenced by its high Cronbach's alphas in Table B.2.

Financial Confidence (FA)

PISA 2018 included a set of questions in `FL162` asking students about their confidence over six financial activities such as making money transfers, understanding bank statements, and plan their spendings using a four-point Likert scale ranging from "Not at all confident", "Not very confident", "Confident" to "Very confident". A variable "confidence about financial matters" was subsequently constructed using the IRT procedure (`FLCONFIN`, OECD (2020c), p. 23). Cronbach's alphas in Table B.2 suggested good reliability.

Financial Application (FB)

The financial literacy application problems were drawn from 43 questions distributed across 24 booklets. The actual test bank remained confidential for reuse, but the OECD was

Table 3.2

Structure of PISA 2018 Financial Literacy Construct

Domain [a]	Content areas	Distribution of score points (%)
Content	Money and transactions	30–40
	Planning and managing finances	25–35
	Risk and reward	15–25
	Financial landscape	10–20
Process	Identify financial information	15–25
	Analyse information in a financial context	15–25
	Evaluate financial issues	25–35
	Applying financial knowledge and understanding	25–35
Contexts	Education and work	10–20
	Home and family	30–40
	Individual	35–45
	Societal	5–15

Note. This table synthesised Table 5.1 to 5.3 of *PISA 2018 Assessment and Analytical Framework* (OECD, 2019a, p. 155). The PISA organiser used the term "score points" instead of "items" because partial credits can be awarded for some questions.

[a] *Content* comprises the areas of knowledge and understanding that are essential in the area of literacy in question; *processes* describes the mental strategies or approaches that are called upon to negotiate the material; and *contexts* refers to the situations in which the knowledge, skills and understandings of the domain are applied, ranging from the personal to the global. (OECD, 2019a, pp. 130–131)

able to provide examples that were comparable in style and difficulty in the *Analytical Framework* (OECD, 2019a, pp. 133–148). These exemplar questions illustrated the domains and content areas (see summary in Table 3.2) PISA 2018 covered for the purpose of constructing candidates' financial literacy scores. In order to succeed in the bank statement question (Figure 5.1, OECD (2019a), p. 133), for example, students should recognise that the necessary information was presented in multiple locations of the financial document and must be identified amongst distractions then summed together. This question covered the "money and transactions" content area of the "content" domain, the "identifying financial information" content area of the "process" domain, and the "home and family" content area of the "contexts" domain. Both constructed- and selected-responses were used in question design and 30 out of 43 items were automatically coded by computers. "Planned missingness" resultant from rotating booklet design was imputed into ten plausible values (von Davier, 2014) centred at 500 with standard deviations of 100 (OECD, 2019a). All ten plausible values (`PV1FLIT` to `PV10FLIT`, collectively written as

FLIT form here on) have been used in subsequent analyses following procedures prescribed by Rubin (1987).

3.2.3 Control Variables

In the 2018 PISA cycle, the OECD simplified its computation of the students' economic, social and cultural status (ESCS) index by taking the arithmetic mean of three indicators: PARED (parental education), HISEI (parental occupational status) and HOMEPOS (home possessions). Figure 16.4 of the *Technical Report* (OECD, 2020c) visualised the ESCS formation procedure while Avvisati (2020) further examined the validity and reliability of the ESCS construct. Students' immigration status was determined by synthesising responses from student questionnaire items ST019 (parents' country of birth) and ST021 (students' age of arrival in test country) (OECD, 2019b, pp. 212–213) into a categorical variable with levels 1 = Native, 2 = Second-Generation and 3 = First-Generation. This information enabled the derivation of two binary variables IMMI1GEN and IMMI2GEN to mark first- and second-generation migrants respectively, with natives being the reference group receiving zero entries for both categories. The variable ST004D01T from the student questionnaire (OECD, 2020a) represented students' gender and was transformed into a binary variable with female being the reference group: 0 = female; 1 = male.

3.3 Multilevel Structural Equation Modelling (MSEM)

Conventional multilevel modelling approaches assume the observed group means to be perfectly reliable when individual-level characteristics are aggregated to the group-level—a particularly questionable assumption in current study. Thanks to recent advancement in both theoretical derivations (Lüdtke et al., 2008; Marsh et al., 2009) and computation power (Muthén & Muthén, 1998–2017), the multilevel latent covariate (MLC) approach has enabled the current project to decompose $L1$ school climate variables FLSCHOOL, FLFAMILY, NOBULLY as well as financial literacy scores FLIT into their corresponding within- and between-level components (subscript $_W$ and $_B$ respectively). This doubly latent MSEM approach controlled measurement error at both the student- and school-levels as well as sampling error due to the aggregation of $L1$ variables to form $L2$ constructs (Lüdtke et al., 2011; Lüdtke et al., 2009; Marsh et al., 2012). Subscript $_{ij}$ in the MSEM model below represents the within-group component of the MLC decomposition and subscript $_j$ stands for the between-group component:

Student-level ($L1$):

$$
\begin{aligned}
\texttt{FCFMLRTY} = \alpha_j^{M_1} &+ \gamma_{11}\texttt{FLSCHOOL}_{ij} + \gamma_{21}\texttt{FLFAMILY}_{ij} + \gamma_{31}\texttt{NOBULLY}_{ij} \\
&+ \gamma_{41}\texttt{ESCS}_{ij} + \gamma_{61}\texttt{IMMI2GEN}_{ij} + \gamma_{71}\texttt{MALE}_{ij} + r_{ij}^{M_1} \\
\texttt{FLCONFIN}_{ij} = \alpha_j^{M_2} &+ \gamma_{12}\texttt{FLSCHOOL}_{ij} + \gamma_{22}\texttt{FLFAMILY}_{ij} + \gamma_{32}\texttt{NOBULLY}_{ij} \\
&+ \gamma_{42}\texttt{ESCS}_{ij} + \gamma_{62}\texttt{IMMI2GEN}_{ij} + \gamma_{72}\texttt{MALE}_{ij} + r_{ij}^{M_2} \\
\texttt{FLIT}_{ij} = \alpha_j^{Y} &+ \beta_1\texttt{FCFMLRTY}_{ij} + \beta_2\texttt{FLCONFIN}_{ij} \\
&+ \gamma_1\texttt{FLSCHOOL}_{ij} + \gamma_2\texttt{FLFAMILY}_{ij} + \gamma_3\texttt{NOBULLY}_{ij} \\
&+ \gamma_4\texttt{ESCS}_{ij} + \gamma_5\texttt{IMMI1GEN}_{ij} + r_{ij}^{Y_W}
\end{aligned}
\tag{3.1}
$$

School-level ($L2$):

$$
\begin{aligned}
\alpha_j^{Y} = \alpha_{00}^{Y} &+ a_1\texttt{FLSCHOOL}_j + a_2\texttt{NOBULLY}_j + a_3\texttt{FLFAMILY}_j + a_4\texttt{EDUSHTG}_j \\
&+ a_5\texttt{STRATIO}_j + \varepsilon_j^{Y_B}
\end{aligned}
\tag{3.2}
$$

with the residual distribution assumptions

$$
\begin{pmatrix} r_{ij}^{M_1} \\ r_{ij}^{M_2} \\ r_{ij}^{Y_W} \end{pmatrix} \sim \mathrm{MVN}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{M_1}^2 & 0 & 0 \\ 0 & \sigma_{M_2}^2 & 0 \\ 0 & 0 & \sigma_{Y_W}^2 \end{pmatrix}\right], \text{ and } \varepsilon_j^{Y_B} \sim \mathcal{N}\left(0,\ \sigma_{Y_B}^2\right), \tag{3.3}
$$

where $\mathrm{MVN}(\cdot)$ and $\mathcal{N}(\cdot)$ stand for multivariate normal and normal distribution respectively.

Using Kaplan's (2009) notation $\boldsymbol{y}_{ij} = \boldsymbol{\alpha}_j + \mathbf{B}_j\boldsymbol{y}_{ij} + \boldsymbol{\Gamma}_j\boldsymbol{x}_{ij} + \boldsymbol{r}_{ij}$ for student-level ($L1$) and random intercept $\boldsymbol{\alpha}_j = \boldsymbol{\alpha}_{00} + \boldsymbol{A}\boldsymbol{w}_j + \boldsymbol{\varepsilon}_j$ for school-level ($L2$), the model equations can be further condensed into the matrix form, with the corresponding path diagram in Figure 3.1:

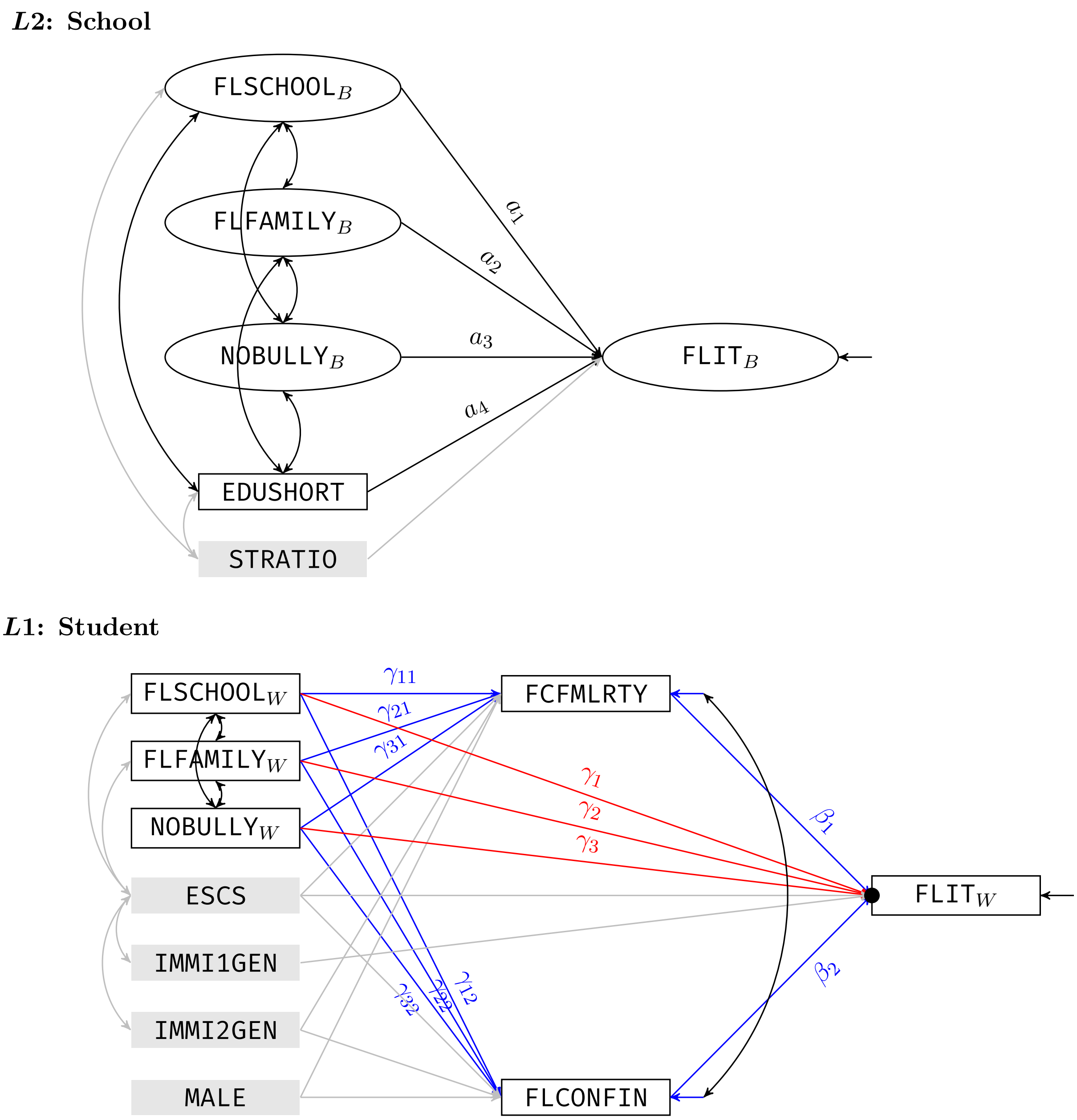
$$
\begin{aligned}
\begin{bmatrix} \texttt{FCFMLRTY}_{ij} \\ \texttt{FLCONFIN}_{ij} \\ \texttt{FLIT}_{ij} \end{bmatrix} &= \begin{pmatrix} \alpha_j^{M_1} \\ \alpha_j^{M_2} \\ \alpha_j^{Y_W} \end{pmatrix} + \begin{pmatrix} 0 & 0 & \beta_1 \\ 0 & 0 & \beta_2 \\ 0 & 0 & 0 \end{pmatrix}^{\intercal} \begin{bmatrix} \texttt{FCFMLRTY}_{ij} \\ \texttt{FLCONFIN}_{ij} \\ \texttt{FLIT}_{ij} \end{bmatrix} \\
&+ \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_1 \\ \gamma_{21} & \gamma_{22} & \gamma_2 \\ \gamma_{31} & \gamma_{32} & \gamma_3 \\ \gamma_{41} & \gamma_{42} & \gamma_4 \\ 0 & 0 & \gamma_5 \\ \gamma_{61} & \gamma_{62} & 0 \\ \gamma_{71} & \gamma_{72} & 0 \end{pmatrix}^{\intercal} \begin{bmatrix} \texttt{FLSCHOOL}_{ij} \\ \texttt{FLFAMILY}_{ij} \\ \texttt{NOBULLY}_{ij} \\ \texttt{ESCS}_{ij} \\ \texttt{IMMI1GEN}_{ij} \\ \texttt{IMMI2GEN}_{ij} \\ \texttt{MALE}_{ij} \end{bmatrix} + \begin{pmatrix} r_{ij}^{M_1} \\ r_{ij}^{M_2} \\ r_{ij}^{Y_W} \end{pmatrix}, \\
\begin{pmatrix} \alpha_j^{M_1} \\ \alpha_j^{M_2} \\ \alpha_j^{Y_W} \end{pmatrix} &= \begin{pmatrix} \alpha_{00}^{M_1} \\ \alpha_{00}^{M_2} \\ \alpha_{00}^{Y} \end{pmatrix} + \begin{pmatrix} 0 & 0 & a_1 \\ 0 & 0 & a_2 \\ 0 & 0 & a_3 \\ 0 & 0 & a_4 \\ 0 & 0 & a_5 \end{pmatrix}^{\intercal} \begin{bmatrix} \texttt{FLSCHOOL}_j \\ \texttt{FLFAMILY}_j \\ \texttt{NOBULLY}_j \\ \texttt{EDUSHTG}_j \\ \texttt{STRATIO}_j \end{bmatrix} + \begin{pmatrix} 0 \\ 0 \\ \varepsilon_j^{Y_B} \end{pmatrix}.
\end{aligned}
\tag{3.4}
$$

3.4 Missing Data Treatment

Missing data are the norm rather than the exception in empirical studies and they demand great care from the researchers to ensure analytical validity. While full information

Figure 3.1

Path Diagram Illustrating the Two-level SEM Predicting Youth's Financial Literacy Outcomes



Note. School climate variables `FLSCHOOL`, `FLFAMILY`, and `NOBULLY`, as well as cognitive outcome `FLIT` are decomposed into the within- and between-components (subscript $_W$ and $_B$ respectively) using the multilevel latent covariate (MLC) approach. Direct pathways are coloured in red while indirect in blue. Control variables are shaded in grey.

maximum likelihood has the benefit of being well understood and readily available in software, the multiple imputation (MI) approach outperforms (a) when the data set contains mixtures of incomplete categorical and continuous variables, (b) when dealing with questionnaire data where items usually come in parcels, (c) when auxiliary variables are required, and (d) when the missing completely at random assumption cannot be reasonably assumed (Enders & Mansolf, 2018). These considerations conclusively directed the current study towards the multilevel MI under the assumption that data were missing at random (Little & Rubin, 2019). In addition, since PISA 2018 financial literacy source files contain missing data at both student- and school-levels and in both continuous and categorical variables, the joint modelling approach is adopted under the advisory of Grund et al. (2018). More specifically, ten sets of imputed data were ordered through **M*plus***'s (Version 8.5, Muthén and Muthén (1998–2017)) unrestricted variance-covariance model ("JM-AM H1", Asparouhov & Muthén, 2010b), using the Bayes estimator with uninformative priors and 4-chain Gibbs sampler to verify convergence as per suggestion by Little and Rubin (2019, p. 230) and Lambert (2018, p. 314). Finally, the first 50,000 burn-in iterations were discarded and any two draws were separated by 5,000 iterations to avoid autocorrelation (see Section B.2.1 for input file)—a safe setting even for moderate to high percentage missings (Grund et al., 2016). See Table B.3 for imputation results and diagnostic plots.

3.5 Sampling Weights

Due to PISA's two-stage sampling design, schools and students were selected with *unequal* probabilities (Chapter 3, OECD (2009), pp. 47–56). A proper incorporation of sampling weights is therefore crucial for establishing unbiased estimations. This study has made use of both student and school weights. Under the advisory of Asparouhov (2006), $L1$ weights were scaled such that they sum to the sample size in each cluster while $L2$ weights were adjusted so that the product of the between- and within-weights sums to the total sample size (Muthén & Muthén, 2017, pp. 622–624).

3.6 Estimator

This study accepted **M*plus***'s default setting of pseudo maximum likelihood (MLR) estimator for the hierarchical modelling (Chapter 16, Muthén & Muthén, 2017, pp. 666 & 668). MLR's robust standard errors are in general Huber-White sandwich estimators (Huber, 1967; White, 1982) with asymptotic standard error corrections using observed residual variances. Literature has long recognised MLR's robust χ^2 tests and standard errors as being more accurate than the asymptotic tests when data are non-normal and when models are mis-specified (Chou et al., 1991; Curran et al., 1996). In the multilevel modelling context, robust χ^2 and standard errors may also provide protection against unmodelled heterogeneity resultant from mis-specification at the group-level or from omitting a level (Hox et al., 2010).

3.7 Model Evaluation

Multiple imputation substantially complicates model fit interpretations. It is important to reflect that Rubin's (1987) rules apply only to *model parameters* under the assumption that over repeated samples, estimates eventually form normal curves peaked at some population values. The distributions of fit indices, on the other hand, are almost always unknown or non-normal, imposing high standards of proof on any proposed aggregation procedures. Early work such as Meng and Rubin (1992) on pooled likelihood-ratio statistic, the precursor to many model fit indices, has been substantiated by simulation studies more recently with encouraging results that it is feasible to construct pooled information criteria (Claeskens & Consentino, 2008) as well as pooled model fit indices (Asparouhov & Muthén, 2010a) under MI. Enders and Mansolf (2018) further suggested that with large samples ($N > 100$) and low missing rates ($< 30\%$–40%), common cut-off criteria such as Hu and Bentler (1999) remain valid. This study took advantage of **M*plus***'s capability of automatically pooling model fit information in the presence of MI. Supported by large sample size ($N = 107,162$) and low missing rate (maximum 22.08%), conventional cut-offs of RMSEA $\leq .06$, SRMR $\leq .08$, CFI $\geq .95$ and TLI $\geq .95$ are likely to be suitable for model comparison purposes.

Iterations whose model fit indices fell short of the abovementioned cut-off criteria were further investigated using modification indices and (fully standardised) expected parameter change (EPC). Modification indices (ModInd) suggest how much a model's χ^2 statistic would decrease by should a fixed parameter were freely estimated; a ModInd greater than 3.84 (critical value of χ^2_1 at $\alpha = .05$) warrants further consideration for theoretical plausibility (Whittaker, 2012). The EPCs, in contrast, indicate the estimated value of a fixed parameter if it were added to a model and freely estimated, providing a more direct estimate of the size of the misspecification for the parameters under consideration. Kaplan (1989) compared ModInd and EPC's impact on empirical studies and concluded that the former had a tendency to suggest freeing implausible parameters while the latter were more likely to recommended reasonable candidates to the model. This study made use of the decision rule prescribed by Saris et al. (1987) to freely estimate a parameter when both ModInd and EPC are large. Model modification decisions were applied sequentially under the advisory of MacCallum et al. (1992) and with close consideration to theoretical ground to ensure underlying substantive assumptions were justified.

Two operational concerns were relevant to the current study. Firstly, since **M*plus*** Version 8.5 only accepts one data set for the modification procedures, the file containing the first plausible value was selected for the model evaluation purposes. Secondly, three versions of the EPC were reported by **M*plus***: E.P.C. (Saris et al., 1987), Std E.P.C (Kaplan, 1989) and StdYX E.P.C. (Chou & Bentler, 1993). This study adopted the latter most version largely due to its invariance property resultant from both parameter and residual standardisations. Improper solutions with standardised estimates greater than 1.0 and/or with negative variances (i.e., Heywood cases) were ignored during decision-making process.

Chapter 4 Results

4.1 Descriptive Statistics and Correlations

Table 4.1 presents descriptive statistics of all measures included in the MSEM model. $L1$ variable `NOBULLY` and $L2$ variable `STRATIO` were highlighted as particularly non-normal due to sizeable disagreements between their means and medians in combination with significant skewness. The MLR estimator introduced in Section 3.6 explicitly takes non-normality into account when computing robust standard errors, safeguard the validity of subsequent analyses. These asymmetric variables suggested that the majority of 15-year-olds experienced safe schools and classrooms overcrowding was uncommon in PISA 2018.

Correlations in Table 4.2 further suggested that schools and families cared about youth's financial literacy in synchrony ($\overline{\rho} \approx .23$) and both efforts were associated with higher cognitive and affective outcomes ($\overline{\rho}$ between .17 and .28). Additionally, students' ESCS were positively correlated with both familiarity with ($\overline{\rho} = .23$) and achievement in ($\overline{\rho} \approx .29$) financial literacy. Lastly, there was a positive correlation between familiarity and confidence ($\overline{\rho} \approx 0.23$) and a similar strength existed between confidence and performance ($\overline{\rho} = 0.23$).

Correlations at the school-level exhibited interesting patterns. Schools with strong emphases on financial education also tended to have engaging parents ($\overline{\rho} \approx .24$), a relationship similar to its $L1$ counterpart in size and magnitude. Although the negative correlation between resource shortage and school safety ($\overline{\rho} \approx -.21$) was expected, it remained counterintuitive that schools that were less safe ($\overline{\rho} \approx -.47$) and were suffering from resource shortages ($\overline{\rho} \approx .31$) tended to be more active in delivering financial education programs. Finally, average performance tended to be higher in safer ($\overline{\rho} \approx .43$) and better equipped ($\overline{\rho} \approx -.44$) schools; while higher levels of school ($\overline{\rho} \approx -.53$) and family interventions ($\overline{\rho} \approx -.36$) have been observed from schools that under-performed in financial literacy.

4.2 Intraclass Correlation and Effective Sample Size

The intraclass correlation ρ_1 can be computed from the random effects ANOVA model ("Null model" in Table 4.3):

Table 4.1

Descriptive Statistics

Analysis level	Variable label	Non-missing sample size	Missing rate (%) [a]	Median	M	SD	Variance	Skewness	Excess kurtosis	Minimum	Maximum
Student (within, $L1$)	`FLSCHOOL`	96435	10.01	0.126	0.018	1.020	1.040	0.189	−0.343	−1.564	2.317
	`FLFAMILY`	95133	11.23	0.011	0.064	1.044	1.090	0.121	0.030	−2.042	2.452
	`NOBULLY`	83499	22.08	0.782	−0.059	1.054	1.110	−1.078	0.664	−3.859	0.782
	`ESCS`	104784	2.22	−0.158	−0.241	1.088	1.183	−0.533	0.184	−7.711	4.234
	`IMMI1GEN`	103317	3.59	0.000	0.029	0.167	0.028	5.608	29.446	0.000	1.000
	`IMMI2GEN`	103317	3.59	0.000	0.042	0.202	0.041	4.542	18.627	0.000	1.000
	`MALE`	107160	0.00	1.000	0.502	0.500	0.250	−0.007	−2.000	0.000	1.000
	`FCFMLRTY`	99969	6.71	7.000	7.049	5.455	29.752	0.223	−1.039	0.000	18.000
	`FLCONFIN`	90130	15.89	−0.027	−0.072	1.017	1.034	−0.084	0.355	−2.210	2.322
	`FLIT` [b]	107162	0.00	481.970	478.291	97.074	9,423.320	−0.089	−0.340	114.256	827.977
School (between, $L2$)	`EDUSHORT`	6346	4.30	0.100	0.131	1.036	1.073	0.341	−0.188	−1.421	2.959
	`STRATIO`	5626	15.16	11.886	13.873	10.171	103.449	4.021	25.425	1.000	100.000

Note. [a] Missing rates were computed based on $N_{L1} = 107,162$ students and $N_{L2} = 6,631$ schools. [b] For descriptive statistics purpose *only*, `FLIT` was obtained by averaging ten plausible values `PV1FLIT` to `PV10FLIT`.

Table 4.2

Correlations between Variables used in the MSEM

	$L1$/within-level	1	2	3	4	5	6	7	8	9	10
1	FLSCHOOL$_W$										
2	FLFAMILY$_W$	$.227^{***}$									
3	NOBULLY$_W$	$-.032^{***}$	$-.044^{***}$								
4	ESCS	$.054^{***}$	$.093^{***}$	$-.003$							
5	IMMI1GEN	$-.002$	$-.001$	$.006$	$.038^{**}$						
6	IMMI2GEN	$-.009$	$.003$	$.019^{\dagger}$	$.040^{*}$	$-.046^{***}$					
7	MALE	$.049^{***}$	$-.039^{***}$	$-.071^{***}$	$.026^{*}$	$-.003$	$-.006$				
8	FCFMLRTY	$.280^{***}$	$.174^{***}$	$.023^{*}$	$.230^{***}$	$-.009$	$-.017$	$.029^{**}$			
9	FLCONFIN	$.201^{***}$	$.190^{***}$	$-.020^{*}$	$.070^{***}$	$.002$	$-.029^{**}$	$.116^{***}$	$.228^{***}$		
10	FLIT$_W$	$-.021^{\dagger}$	$.021^{*}$	$.053^{***}$	$.288^{***}$	$-.029^{*}$	$.025^{\dagger}$	$.020^{\dagger}$	$.230^{***}$	$.068^{***}$	

	$L2$/between-level	11	12	13	14	15	16
11	FLSCHOOL$_B$						
12	FLFAMILY$_B$	$.239^{**}$					
13	NOBULLY$_B$	$-.468^{***}$	$-.065$				
14	EDUSHORT	$.313^{***}$	$.053$	$-.207^{**}$			
15	STRATIO	$-.082^{*}$	$.131^{*}$	$.026$	$-.043$		
16	FLIT$_B$	$-.529^{***}$	$-.356^{***}$	$.426^{***}$	$-.438^{***}$	$-.101^{**}$	

Note. The MLC procedure decomposes school climate variables FLSCHOOL, FLFAMILY and NOBULLY as well as financial literacy outcomes FLIT into their within- and between-components (subscript $_W$ and $_B$ respectively). Correlations at each level refer to the maximum-likelihood estimated within- and between-covariance matrices respectively. All statistics are average results over ten imputed data sets, denoted as $\overline{\rho}$ in the text.

$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

$$\rho_1 = \frac{\text{School-level residual variance}}{\text{Total residual variance}} = \frac{\text{var}\left(\varepsilon_j^{Y_B}\right)}{\text{var}\left(r_{ij}^{Y_W}\right) + \text{var}\left(\varepsilon_j^{Y_B}\right)} = \frac{5240}{6122 + 5240} = 0.461. \quad (4.1)$$

This result suggested that 46.1% of the variation in financial literacy performance was due to the clustering in schools.

For sample size adjustment, Snijders and Bosker (2012) advised to first of all calculate the design effect of one's multilevel model:

$$\text{design effect} = 1 + (\text{average group size} - 1)\rho_1 = 1 + \left(\frac{107,162}{6,631} - 1\right) \times 0.461 = 7.989, \quad (4.2)$$

then compute the effective sample size:

$$N_{\text{effective}} = \frac{N_{\text{original}}}{\text{design effect}} = \frac{107,162}{7.989} = 13,414. \quad (4.3)$$

This result signaled that students from the same school were so similar in their financial literacy outcomes that the sample size of 107,162 used by this study was equivalent to a simple random sample using 13,414 students. This result not only provided assurance of a sufficiently large sample size required by asymptotic theories but also highlighted the strong effect of schools for understanding youth's financial literacy development.

4.3 Intermediate Models

In order to separate the incremental effect attributable to school-level variables, a student-level only model was first established as a reference ("Single-level model" in Table 4.3). Even with $L1$-only variables, model fit indices CFI = .97, TLI = .927 and SRMR = .016 jointly suggested that the proposed input (school climate)–mediator (FC & FA)–output (FB) model was a meaningful one. Next, school-level variables were allowed to covary between one other on top of the $L1$ structure, forming a two-level saturated model. This procedure had an effect of decomposing the total residual variances into student- and school-levels. As a result, $L1$ residual variance reduced by more than a quarter from 7,866 to 5,764, indicating the necessity of the $L2$ structure.

4.4 Full Model

Relationships amongst school-level variables were further introduced at $L2$, transforming the saturated model into the final MSEM model illustrated in Figure 3.1.

4.4.1 Model Fit

Model fit indices CFI = .968, SRMR_{L1} = .015 and SRMR_{L2} = .030 all satisfied the cut-off criteria suggested by Hu and Bentler (1999) while TLI = .903 fell slightly short of being good but still acceptable—a penalty on the growing number of variables introduced. On balance, there was sufficient evidence suggesting good fit between the proposed MSEM model and financial literacy data.

4.4.2 Student-level Relationships

School Climate Variables

All three $L1$ school climate variables shared statistically significant relationships with financial literacy performance (`FLIT`). A safe school environment (`NOBULLY`) was positively correlated with financial literacy via both the direct pathway and through mediation with familiarity (`FCFMLRTY`).

Efforts by schools (`FLSCHOOL`) and families (`FLFAMILY`), on the other hand, had more nuanced relationships with the cognitive outcome. Both variables had strong positive associations with `FLIT` via mediation pathways, but statistically significant *negative* relationships via direct pathways. Such positive-negative pair happened to cancel each other for `FLFAMILY`, leading to a non-significant result should financial socialisation and financial literacy were correlated superficially. The negative cognitive path overshadowed the positive affective pathways for `FLSCHOOL`, leading to a seemingly paradoxical negative overall relationship between classroom efforts and financial literacy scores.

Demographic Attributes

The strongest covariation identified by this study was between students' ESCS and their financial literacy outcomes. Substantial positive associations have been observed along both the direct and indirect pathways. Having controlled ESCS as a confounder is therefore essential for the study of school climate effects.

The relationship between one's immigration history and their financial literacy performance also delivered important insight. Children who relocated to the host country between births and reaching 15-year-old (`IMMI1GEN` = 1) seemed to possess less application skills in financial matters whereas the offspring of migrants did not show deficiency via knowledge and confidence.

Meanwhile, school curricula addressing students' affinity towards finance-related topics would likely to benefit not only second-generation migrants but also young girls. This conjecture was made based on the observed male advantage in financial literacy performance—everything

Table 4.3

Model Parameters and Fit Indices of Multilevel Regressions for the Global Sample

Variable — path	Model parameter	Null model Coef	*SE*	Single-level model Coef	*SE*	Two-level saturated Coef	*SE*	Two-level structured Coef	*SE*
FIXED EFFECTS									
Intercept		454.154	2.690***	451.451	1.449***	445.812	2.578***	486.820	4.500***
Student-level Predictors									
`FLSCHOOL` (total)	$\gamma_1 + \gamma_{11}\beta_1 + \gamma_{12}\beta_2$			−0.073	0.008***	−0.036	0.011**	−0.036	0.011**
— direct	γ_1			−0.125	0.008***	−0.088	0.011***	−0.088	0.011***
— total indirect	$\gamma_{11}\beta_1 + \gamma_{12}\beta_2$			0.051	0.003***	0.052	0.003***	0.052	0.003***
— via `FCFMLRTY`	$\gamma_{11}\beta_1$			0.049	0.002***	0.047	0.003***	0.047	0.003***
— via `FLCONFIN`	$\gamma_{12}\beta_2$			0.002	0.001	0.005	0.002**	0.005	0.002**
`FLFAMILY` (total)	$\gamma_2 + \gamma_{21}\beta_1 + \gamma_{22}\beta_2$			0.008	0.007	0.005	0.009	0.005	0.009
— direct	γ_3			−0.016	0.007*	−0.019	0.009*	−0.019	0.009*
— total indirect	$\gamma_{21}\beta_1 + \gamma_{22}\beta_2$			0.023	0.002***	0.024	0.002***	0.024	0.002***
— via `FCFMLRTY`	$\gamma_{21}\beta_1$			0.022	0.002***	0.019	0.002***	0.019	0.002***
— via `FLCONFIN`	$\gamma_{22}\beta_2$			0.002	0.001	0.005	0.002**	0.005	0.002***
`NOBULLY` (total)	$\gamma_3 + \gamma_{31}\beta_1 + \gamma_{32}\beta_2$			0.075	0.007***	0.053	0.009***	0.053	0.009***
— direct	γ_3			0.064	0.007***	0.046	0.009***	0.046	0.009***
— total indirect	$\gamma_{31}\beta_1 + \gamma_{32}\beta_2$			0.011	0.002***	0.007	0.002***	0.007	0.002***
— via `FCFMLRTY`	$\gamma_{31}\beta_1$			0.011	0.002***	0.007	0.002***	0.007	0.002***
— via `FLCONFIN`	$\gamma_{32}\beta_2$			0.000	0.000	0.000	0.000	0.000	0.000
`ESCS` (total)	$\gamma_4 + \gamma_{41}\beta_1 + \gamma_{42}\beta_2$			0.497	0.007***	0.289	0.016***	0.289	0.016***
— direct	γ_4			0.445	0.007***	0.248	0.015***	0.248	0.015***
— total indirect	$\gamma_{41}\beta_1 + \gamma_{42}\beta_2$			0.052	0.003***	0.041	0.003***	0.041	0.003***
— via `FCFMLRTY`	$\gamma_{41}\beta_1$			0.052	0.002***	0.040	0.003***	0.040	0.003***
— via `FLCONFIN`	$\gamma_{42}\beta_2$			0.001	0.001	0.001	0.001*	0.001	0.001*
`IMMI1GEN` (direct)	γ_5			0.004	0.008	−0.040	0.012**	−0.040	0.012**
`IMMI2GEN` (total indirect)	$\gamma_{61}\beta_1 + \gamma_{62}\beta_2$			−0.003	0.002†	−0.006	0.002**	−0.006	0.002**
— via `FCFMLRTY`	$\gamma_{61}\beta_1$			−0.003	0.002†	−0.005	0.002*	−0.005	0.002*
— via `FLCONFIN`	$\gamma_{62}\beta_2$			0.000	0.000	−0.001	0.000*	−0.001	0.000*
`MALE` (total indirect)	$\gamma_{71}\beta_1 + \gamma_{72}\beta_2$			0.004	0.002*	0.007	0.002**	0.007	0.002**
— via `FCFMLRTY`	$\gamma_{71}\beta_1$			0.003	0.002*	0.004	0.002*	0.004	0.002*
— via `FLCONFIN`	$\gamma_{72}\beta_2$			0.001	0.001	0.003	0.001**	0.003	0.001**

Continued

Variable	Model parameter	Null model Coef	Null model *SE*	Single-level model Coef	Single-level model *SE*	Two-level saturated Coef	Two-level saturated *SE*	Two-level structured Coef	Two-level structured *SE*
School-level Predictors									
`FLSCHOOL`	a_1							-0.295	0.066***
`FLFAMILY`	a_2							-0.225	0.057***
`NOBULLY`	a_3							0.233	0.069**
`EDUSHORT`	a_4							-0.292	0.038***
`STRADIO`	a_5							-0.132	0.026***
RANDOM EFFECTS (residual variances of `FLIT`)									
Student-level	$\text{var}\left(r_{ij}^{Y_W}\right)$	6121.904	131.192	7866.408	114.555	5763.677	130.133	5763.690	130.133
School-level	$\text{var}\left(\varepsilon_{j}^{Y_B}\right)$	5240.477	202.004			3264.618	193.892	1705.616	135.044
MODEL FIT INDICES		Est	*SD*	Est	*SD*	Est	*SD*	Est	*SD*
AIC		1253984	1093	3429058	1534	3468075	1661	3468108	1650
BIC		1254013	1093	3429566	1534	3468727	1661	3468740	1650
χ^2 Test of Model Fit		2.193	1.468	304.405	13.167	187.655	10.486	201.645	11.746
RMSEA		0.000	0.000	0.017	0.000	0.009	0.000	0.009	0.000
CFI		0.000	.000	.970	.002	.970	.002	.968	.002
TLI		1.000	.000	.927	.004	.899	.007	.903	.007
SRMR $L1$		.005	.003	.016	.000	.015	.000	.015	.000
SRMR $L2$		.011	.005			.014	.002	.030	.006

Note. All p values in this table are two-tailed.

$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

else being equal, 15-year-old boys on average demonstrated higher financial capability, a fully mediated effect particularly through higher confidence.

4.4.3 School-level Relationships

Shortages in either capital or labour resources were associated with lower average financial literacy outcomes at the school-level. The MSEM showed a negative relationship between the fourth element of school climate variable, educational resource shortage `EDUSHORT`, and average `FLIT`. In fact, the association between schools' physical capital and their educational output remained one of the strongest statistical relationships identified by this study, over twice the size of that between labour arrangement (student-teacher ratio `STRATIO`) and financial literacy achievement.

4.4.4 Contextual Effects

One particular strength of an MSEM is its ability to model contextual effects. In a school research context, there exists a *contextual effect* when school-level characteristics contribute to individual learners' outcomes beyond what can be explained by student-level characteristics. Following Marsh et al. (2009)'s procedure, this study obtained the point estimate of the unstandardised contextual effect for `FLSCHOOL`:

$$\text{Unstandardised contextual effect} = \widehat{a}_1 - \widehat{\gamma}_1 = -49.339 - (-7.078) = -42.261, \tag{4.4}$$

and its standardised solution:

$$\begin{aligned} &\text{Standardised contextual effect} \\ &= \frac{\text{Unstandardised contextual effect} \times \sqrt{\widehat{\text{var}}\left(\texttt{FLSCHOOL}_B\right)}}{\sqrt{\widehat{a}_1^2 \cdot \widehat{\text{var}}\left(\texttt{FLSCHOOL}_B\right) + \widehat{\text{var}}\left(\texttt{FLIT}_B\right) + \widehat{\gamma}_1^2 \cdot \widehat{\text{var}}\left(\texttt{FLSCHOOL}_W\right) + \widehat{\text{var}}\left(\texttt{FLIT}_W\right)}} \\ &= \frac{(-42.261) \times \sqrt{0.114}}{\sqrt{(-49.339)^2 \times 0.114 + 3226.753 + (-7.078)^2 \times 1.009 + 6576.975}} \\ &= -\,0.163,\ \ (-0.142 \text{ if calculated manually due to cumulative rounding errors}) \end{aligned} \tag{4.5}$$

while the associated standard error can be obtained using the delta method (Raykov & Marcoulides, 2004). Table 4.4 summarised the contextual effect estimates for `FLSCHOOL`, `FLFAMILY`, and `NOBULLY`. These results suggested that students' financial literacy performance was not only affected by individual characteristics and endeavour but also heavily influenced by the larger school environment surrounding the learners. Lastly, the effect size (ES) statistics in Table 4.4 further suggested that the significant contextual effect findings were unlikely to be a mere statistical artefact out of large sample sizes, evidenced by their large sizes ($|\text{ES}| \approx .38$ and .33) and robustness against various of calculation methods (conventional ES1 by Tymms (2004) and recent innovations ES2 and ES3 by Marsh et al. (2009)).

Table 4.4

Contextual Effects and Effect Sizes

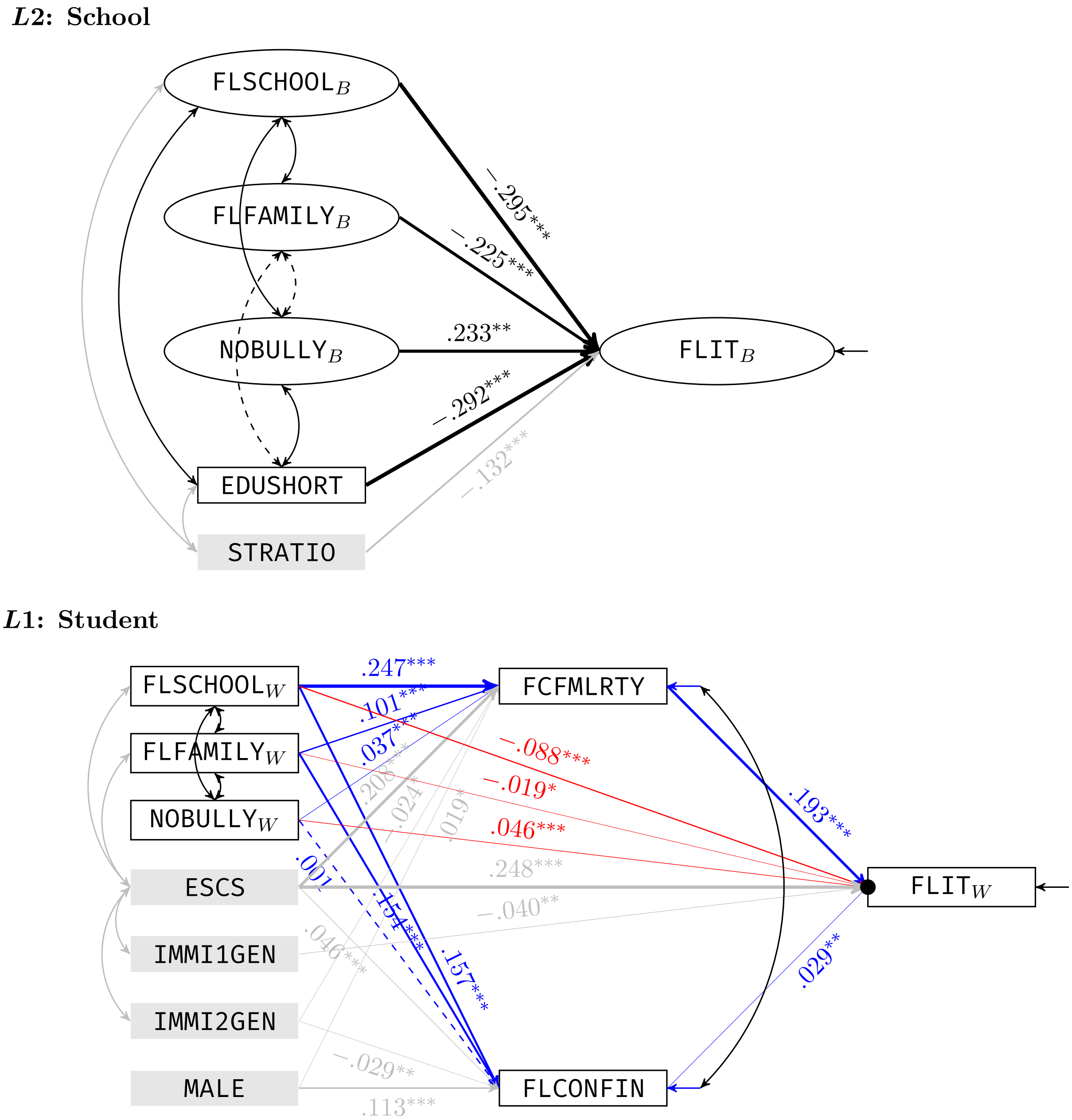
Contextual relationship	Contextual effect Est	Contextual effect *SE*	**Standardised contextual effect** Est	**Standardised contextual effect** *SE*
FLSCHOOL	−42.261	10.720***	−0.163	0.041***
FLFAMILY	−75.808	20.353***	−0.144	0.037***
NOBULLY	60.071	19.673**	0.144	0.046**

Contextual relationship	Effect size 1 Est	Effect size 1 *SE*	**Effect size 2** Est	**Effect size 2** *SE*	Effect size 3 Est	Effect size 3 *SE*
FLSCHOOL	−0.380	0.099***	−0.378	0.098***	−0.369	0.092***
FLFAMILY	−0.332	0.084***	−0.332	0.084***	−0.328	0.081***
NOBULLY	0.331	0.107**	0.331	0.107**	0.326	0.102**

Note. Contextual effect computations and standardisations were based on the procedure documented in Marsh et al.'s (2009) supplemental Model 8. Marked in bold, standardised contextual effect and effect size 2 were recommended for decision-making. Effect sizes 1 (Tymms, 2004) was provided as reference due to its compatibility with Cohen's d (Cohen, 1992). More recently, Marsh et al. (2009) advocated for an effect size procedure involving total variances from *both* levels (ES3) over that from only $L1$ (ES2) (see Marsh et al., 2009, p. 792). Since consensus so far appears to be with ES2, ES3 was provided for future reference.

Figure 4.1

Two-level Structural Equation Model Predicting Youth's Financial Literacy Outcomes



Note. This multilevel structural equation model predicts youth's financial literacy outcomes using PISA 2018 data, with mediating effects of familiarity with, and confidence in, financial matters at student-level. Statistics are standardised regression coefficients. Dashed lines represent nonsignificant relations at $\alpha = .05$ level. Student-level school climate variables and cognitive outcome are decomposed into the within- and between-components (subscript $_W$ and $_B$ respectively) using the MLC approach. Direct pathways are coloured in red and indirect in blue. Control variables are shaded in grey.

$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Chapter 5 Discussion

5.1 Overview

"It takes a village to raise a child." This study looked into the dual mechanisms of how factors associated with 15-year-old students' financial literacy related to each other (RQ 1) and how the surrounding school environment may facilitate such relationships (RQ 2). MSEM results showed that 33.5% of the variation in students' `FLIT` scores can be explained by student-level variables and 47.7% by school-level factors (see Section B.3.2), suggesting the importance of schools in cultivating youth's financial literacy outcomes. By accounting for the hierarchical data structure, sampling weights, missing data imputation, as well as measurement error and sampling error, this study was able to ascertain the marginal effects of the four school climate variables: ACADEMIC, COMMUNITY, SAFETY and INSTITUTIONAL ENVIRONMENT (Wang & Degol, 2016) respectively (see Figure 4.1 and Table 4.3). This study added empirical evidence to Kutsyuruba et al.'s (2015) review article by showing the importance of school safety for students' financial knowledge, confidence, and application behaviour. The student-level model extended Jorgensen and Savla's (2010) structural equation approach to financial literacy and confirmed the key roles financial knowledge ($R^2 = .136$) and confidence ($R^2 = .077$) played in mediating youth's financial literacy achievement.

This study also revealed a key insight that was initially less intuitive. At both individual- and school-levels, the associations between explicit teaching of finance-related topics (FEdu) and contemporaneous financial literacy performance (FB) were found to be *negative*. In addition, the relationships between parental involvement (FSoc) for cultivating youth's financial literacy outcomes were shown to be positive along the mediation pathways (via FC and FA) but negative along the application pathway (FB). These two effects were similar in size but opposite in sign. At the school-level, both classroom activities and parental care, on average, tended to be more visible around students who were yet to demonstrate their mastery of financial capabilities. Sizeable contextual effects further suggested schools rather than learners as the source of the observed negative correlations between financial literacy outcome (FB) and teaching efforts (FEdu), and between FB and financial socialisation (FSoc).

5.2 Responses to the Research Questions and Hypotheses

5.2.1 Research Question 1

All four school climate variables explained variation in youth's financial literacy outcomes. Financial knowledge (FC) and confidence (FA) played significant mediation roles for explaining financial literacy scores (FB), confirming Hypotheses 12 and 13. This result was partially consistent with Jorgensen and Savla's (2010) mediation model in which the author focused solely on the relationships between parental influence and financial behaviour, mediated by financial knowledge and attitude. This study effectively corrected Jorgensen and Savla's (2010) omitted variable problem by adding back FEdu, Safety and more demographic controls at $L1$ and an additional structure at $L2$, subsequently re-establishing FC as a significant mediator. Such result was fully expected under the family financial socialisation theory (Danes & Haberman, 2007) where financial knowledge development shall be an important component.

Financial education (FEdu) showed positive effects along the mediation pathways (confirming H1, H2) but a negative effect along the direct pathway (contradicting H3) with financial literacy scores (FB). Since the direct effect overshadowed the mediation effects, the total effect between FEdu and FB appeared to be negative. This result positioned the current study in line with a series of papers reporting non-significant or negative findings. Studies using the test-retest design (Mandell & Klein, 2009), randomised experiment with treatment-control groups (Becchetti et al., 2013; Collins, 2013) as well as an archival study using PISA 2012 data (Farinella et al., 2017) all questioned the effectiveness of financial education courses. Additionally, Mountain et al.'s (2020) 5-year-horizon longitudinal study identified a negative association between long-term financial behaviours and attending workshops and seminars, mediated by financial knowledge. In light of these publications, the negative direct pathway identified by the current study shall not be dismissed as an statistical irregularity but an invitation for further considerations (see Section 5.3).

Similar to FEdu, Parental involvement at home (FSoc) had positive mediation pathways (confirming H4 and H5) but an equi-magnitude negative direct pathway (contradicting H6), leading to a non-significant total relationship between FSoc and FB. This result shall be differentiated from the positive FSoc-FB association by Moreno-Herrero et al. (2018) since the latter design did not involve FEdu, Safety or any mediators, leading to a possible redistribution of explanatory power from the omitted variables into FSoc.

Safety was found to have positive effects for students' financial knowledge, confidence, as well as application behaviour (confirming H7, H8 and H9), linking Kutsyuruba et al.'s (2015) school safety review to the financial literacy research.

5.2.2 Research Question 2

All four school climate variables at the school-level were shown to be statistically significant for explaining the variation in school-average financial literacy scores. MSEM results revealed that educational resource shortages as well as high student-teacher ratios both correlated with lower average financial literacy performance, confirming H10 and H11 and the applicability of prior studies (Finn & Achilles, 1999; Miles & Darling-Hammond, 1998; Uline & Tschannen-Moran, 2008) to the field of financial literacy research.

Adding to existing literature, FEdu, FSoc and Safety were all shown to have significant contextual effects, suggesting individual students' financial literacy capability was strongly affected by their school environment. Along with the higher R^2 observed at $L2$ (see Section B.3.2), and the strong design effect calculated in Equation (4.2), the current study consistently highlighted school-level factors as the driving force behind the systematic variations in students' PISA 2018 financial literacy performance.

5.3 Conjectures about Negative Pathways

Although causal inferences could not be established from a correlational study design, a negative association between input and output variables may still *suggest* some interesting possibilities for future studies. If one hypothesises a causal direction $\texttt{FLSCHOOL} \longrightarrow \texttt{FLIT}$, the negative relationship between the two variables could signal potential improvement opportunities for current financial education practices. While students have benefited from educational interventions with growing knowledge and confidence, existing pedagogy may yet to explicitly train students to link their learning to real-world finance problem-solving. Bridging the disconnect between minds and hands has long been emphasised in science (Harlen, 1999) and mathematics (Smith et al., 1996) education and voices for learning from sister subjects' success started to grow in the field of financial education (Marley-Payne et al., 2021). Parents may similarly adapt by introducing financial problem-solving skills in addition to sharing knowledge and affects at home. Alternatively, a causal direction $\texttt{FLSCHOOL} \longleftarrow \texttt{FLIT}$ may suggest that educational and parental attention was being directed preferentially towards students who were most in need of developing problem-solving skills—it was not the quality of interventional efforts but the insufficient quantity that needed to be addressed. Future research may investigate the plausibility of such constraint optimisation behaviour by teachers and parents and estimate the sizes of the Lagrange multipliers as evidence for the potential marginal improvement should schooling and parenting resources were expanded. A third possibility involves a hidden confounder $\texttt{FLSCHOOL} \longleftarrow \text{confound} \longrightarrow \texttt{FLIT}$. Jappelli's (2010) observation that students' financial literacy tended to be lower in countries with stronger social safety net could serve as a starting point for this line of investigation under the reasonable assumption that such countries also devote higher social resources into education input. Should this direction of study become fruitful, financial educators would then be reminded the importance of social arrangement as a moderator, where it would be desirable to re-allocate educational resources taking into account each society's social contracts.

A non-linear relationship could be a fourth possibility for the negative association between FLSCHOOL and FLIT. Using 2015 TIMSS data, Teig et al. (2018) demonstrated a curvilinear relationship between inquiry-based teaching practice and students' science achievement with high frequency inquiry-based teaching being linked to a reduced performance. A quadratic relationship was reported between learning time and science achievement using PISA 2015 data (Zhang et al., 2021) especially in Eastern cultures, possibly indicating that non-linearity could become a relative common consideration when analysing large-scale international assessment data. A verification of similar curvilinear relationship in the financial literacy filed is important so that educational and parental resources can be further optimised.

A final hypothesis can be made based on the implementation lags observed by Bernheim et al. (2001). Financial literacy could be unique in a sense that it requires a longer time for FEdu and FSoc to be consolidated, incorporated and then turned into observable behaviour improvement, including application and problem-solving behaviour. That is to say the negative relationship between $\texttt{FLSCHOOL}_t$ and $\texttt{FLIT}_{t-1}$ reflected the maturing effect of financial skill acquisition process. A longitudinal study is required in order to confirm this intertemporal growth model.

5.4 Limitations

The correlational research design used by this study limited the possible causal inferences. Using Shadish et al.'s (2002) taxonomy, this study demonstrated strong statistical conclusion validity by showing both the presence and strength of the covariation between school climate variables and students' financial literacy outcomes. It was unable to, however, demonstrate whether school climate preceded financial literacy in time, neither was it able to exclude all other relationships as plausible explanations for the covariation between the two. By this measure, the current study's internal validity is not yet strong. As the scholarly world is yet to reach consensus on the best construct to represent financial literacy, this study inherited one particular version of financial literacy operationalised by the PISA organiser, whose construct validity continues to attract scrutiny by both theorists and practitioners (Schuhen & Schürkmann, 2014). Lastly, statistical parameters derived in this study were based on data drawn from predominantly industrialised countries, questioning its strength on external validity.

The other limitation originated from the data design. Since this study pooled all 20 participating countries into a global data structure, the subsequent analyses and statistical results must be interpreted as the global, rather than country-specific outcomes. This observation is important for education policy making since global averages may not serve the interests of local conditions correctly. Since industrialised economies were over-represented in the 20-country sample, pedagogical and policy implications may be skewed towards countries with similar socio-economic profiles. Further studies are encouraged to replicate procedures employed by this project by counties in order to obtain evidence better situated with the local environment.

Based on the limitations discussed above, future research efforts may consider upgrading the study design from a correlational to a causal one by using, amongst others, instrumental variable (Pokropek, 2016) or panel data (Salas-Velasco, 2019) techniques. Country-by-country comparisons would also provide additional insight into the similarities and differences across economies, aiding pedagogy design and education policy formation processes.

5.5 Contribution and Conclusions

This research project contribute to financial literacy literature in a number of ways. It first of all linked a substantive theoretical framework of school climate to youth's financial literacy development process in order to examine how individuals' capability is formed *in the context of* their school environment. This person-ecological approach reconciled two strands of research efforts that focused either on students or on schools into one unified structure. In terms of methodology, this study attempted a recent development in the MSEM literature using a multilevel latent covariate approach (MLC, Lüdtke et al. (2008); "doubly-latent model", Marsh et al. (2009)) to correct for unreliability at higher-level when lower-level constructs were aggregated up. The successful application of this new technique to the most recent PISA 2018 data set showcased the advancement in the field of educational measurement.

A well-functioning society relies on citizens' financial literacy for the betterment of their own well-being and that of the collective. Policy-makers, school leaders, teachers and parents all have progressively come to terms with the cost of neglect and demanded evidence-based action plans. The current research project answered this call by exploring four aspects of school climate using the latest international large-scale assessment data—Education matters. Parenting matters. Safety and resource fundings do matter. These conclusions shed light to the policy priorities that can be actioned upon without delay. This study served only as a starting point for a vibrant scholarly conversation about better preparing our young for an ever-challenging future. May they benefit and succeed.

References

Agarwal, S., Amromin, G., Ben-David, I., Chomsisengphet, S. & Evanoff, D. D. (2015). Financial literacy and financial planning: Evidence from India. *Journal of Housing Economics*, *27*, 4–21. https://doi.org/10.1016/j.jhe.2015.02.003

Agarwal, S., Chomsisengphet, S. & Zhang, Y. (2015). How does financial literacy affect mortgage default? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2601025

Agnew, S. & Cameron-Agnew, T. (2015). The influence of consumer socialisation in the home on gender differences in financial literacy. *International Journal of Consumer Studies*, *39*(6), 630–638. https://doi.org/10.1111/ijcs.12179

Agnew, S. & Harrison, N. (2015). Financial literacy and student attitudes to debt: A cross national study examining the influence of gender on personal finance concepts. *Journal of Retailing and Consumer Services*, *25*, 122–129. https://doi.org/10.1016/j.jretconser.2015.04.006

Akben-Selcuk, E. & Altiok-Yilmaz, A. (2014). Financial literacy among Turkish college students: The role of formal education, learning approaches, and parental teaching. *Psychological Reports*, *115*(2), 351–371. https://doi.org/10.2466/31.11.pr0.115c18z3

Ali, P., Anderson, M., McRae, C. & Ramsay, I. (2016). The financial literacy of young people: Socio-economic status, language background, and the rural-urban chasm. *Australian & International Journal of Rural Education*, *26*(1), 53–65. http://web.a.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=2&sid=f7900e91-7c14-4983-9c5e-41979ed77f10@sdc-v-sessmgr03

Ali, P., Anderson, M. E., McRae, C. H. & Ramsay, I. (2014). The financial literacy of young Australians: An empirical study and implications for consumer protection and ASIC's National Financial Literacy Strategy. *Company and Securities Law Journal*, *32*(5), 334–352. https://ssrn.com/abstract=2490154

Amagir, A., Groot, W., van den Brink, H. M. & Wilschut, A. (2018). A review of financial-literacy education programs for children and adolescents. *Citizenship, Social and Economics Education*, *17*(1), 56–80. https://doi.org/10.1177/2047173417719555

Arceo-Gómez, E. O. & Villagómez, F. A. (2017). Financial literacy among Mexican high school teenagers. *International Review of Economics Education*, *24*, 1–17. https://doi.org/10.1016/j.iree.2016.10.001

Arellano, A., Cámara, N. & Tuesta, D. (2014). *The effect of self-confidence on financial literacy* (Working Paper No. 14/28). BBVA Research. https://www.bbvaresearch.com/wp-content/uploads/2014/10/WP14-28The-effect-of-self-confidence-on-financial-literacy1.pdf

Arthur, C. (2016). Financial literacy education as a public pedagogy: Consumerizing economic insecurity, ethics and democracy. In C. Aprea, E. Wuttke, K. Breuer, N. K. Koh, P. Davies, B. Greimel-Fuhrmann & J. S. Lopus (Eds.), *International handbook of financial literacy* (pp. 113–125). Springer. https://doi.org/10.1007/978-981-10-0360-8_9

Asparouhov, T. (2006). General multi-level modeling with sampling weights. *Communications in Statistics — Theory and Methods*, *35*, 439–460. https://doi.org/10.1080/03610920500476598

Asparouhov, T. & Muthén, B. (2010a). *Chi-square statistics with multiple imputation* (Version 2). Muthén & Muthén. https://www.statmodel.com/download/MI7.pdf

Asparouhov, T. & Muthén, B. (2010b). *Multiple imputation with Mplus* (Version 2). Muthén & Muthén. https://www.statmodel.com/download/Imputations7.pdf

Atkinson, A. & Messy, F.-A. (2011). Assessing financial literacy in 12 countries: An OECD/INFE international pilot exercise. *Journal of Pension Economics and Finance*, *10*(4), 657–665. https://doi.org/10.1017/s1474747211000539

Avvisati, F. (2020). The measure of socio-economic status in PISA: A review and some suggested improvements. *Large-scale Assessments in Education*, *8*(8), 1–37. https://doi.org/10.1186/s40536-020-00086-x

Barnard, C. R., Billing, J., Brotherston, D., Jeffery, T., Mansell, P. & Wright, J. (2021). Money, knowledge and power. *British Actuarial Journal*, *26*(e4), 1–26. https://doi.org/10.1017/s1357321721000039

Bay, C., Catasús, B. & Johed, G. (2014). Situating financial literacy. *Critical Perspectives on Accounting*, *25*(1), 36–45. https://doi.org/10.1016/j.cpa.2012.11.011

Becchetti, L., Caiazza, S. & Coviello, D. (2013). Financial education and investment attitudes in high schools: Evidence from a randomized experiment. *Applied Financial Economics*, *23*(10), 817–836. https://doi.org/10.1080/09603107.2013.767977

De Beckker, K., De Witte, K. & Van Campenhout, G. (2019). Identifying financially illiterate groups: An international comparison. *International Journal of Consumer Studies*, *43*(5), 490–501. https://doi.org/10.1111/ijcs.12534

Belás, J., Nguyen, A., Smrčka, L., Kolembus, J. & Cipovová, E. (2016). Financial literacy of secondary school students. Case study from the Czech Republic and Slovakia. *Economics & Sociology*, *9*(4), 191–206. https://doi.org/10.14254/2071-789x.2016/9-4/12

Bernheim, B., Garrett, D. M. & Maki, D. M. (2001). Education and saving: The long-term effects of high school financial curriculum mandates. *Journal of Public Economics*, *80*(3), 435–465. https://doi.org/10.1016/s0047-2727(00)00120-1

Birbili, M. & Kontopoulou, M. (2015). Financial education for preschoolers: Preparing young children for the 21st Century. *Childhood Education*, *91*(1), 46–53. https://doi.org/10.1080/00094056.2015.1001670

Böhm, P., Böhmová, G., Šimková, V. & Gazdíková, J. (2021). The impact of secondary education on the level of financial literacy: The case of Slovakia. *Problems of Education in the 21st Century*, *79*(1), 13–33. https://doi.org/10.33225/pec/21.79.13

Boisclair, D., Lusardi, A. & Michaud, P.-C. (2017). Financial literacy and retirement planning in Canada. *Journal of Pension Economics and Finance*, *16*(3), 277–296. https://doi.org/10.1017/s1474747215000311

Borges Ramalho, T. & Forte, D. (2019). Financial literacy in Brazil—Do knowledge and self-confidence relate with behavior? *RAUSP Management Journal*, *54*(1), 77–95. https://doi.org/10.1108/RAUSP-04-2018-0008

Bottazzi, L. & Lusardi, A. (2020). *Stereotypes in financial literacy: Evidence from PISA* (NBER Working Paper No. 28065). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w28065/w28065.pdf

Bover, O., Hospido, L. & Villanueva, E. (2018). *The impact of high school financial education on financial knowledge and saving choices: Evidence from a randomized trial in Spain* (Working paper No. 1801). Banco de España. https://doi.org/10.2139/ssrn.3116054

Bowen, C. F. (2002). Financial knowledge of teens and their parents. *Journal of Financial Counseling and Planning*, *13*(2), 93–102. https://afcpe.buckeyedev.com/wp-content/uploads/2018/10/vol1328.pdf

Breitbach, E. & Walstad, W. B. (2016). Financial literacy and financial behavior among young adults in the United States. In E. Wuttke, J. Seifried & S. Schumann (Eds.), *Economic* *competence and financial literacy of young adults* (pp. 81–98). Verlag Barbara Budrich. https://doi.org/10.2307/j.ctvbkk29d.7

Brown, M., Grigsby, J., van der Klaauw, W., Wen, J. & Zafar, B. (2016). Financial education and the debt behavior of the young. *Review of Financial Studies*, *29*(9), 2490–2522. https://doi.org/10.1093/rfs/hhw006

Bucciol, A. & Veronesi, M. (2014). Teaching children to save: What is the best strategy for lifetime savings? *Journal of Economic Psychology*, *45*, 1–17. https://doi.org/10.1016/j.joep.2014.07.003

Bucher-Koenen, T., Alessie, R., Lusardi, A. & van Rooij, M. (2021). *Fearless girl: Woman's financial literacy an stock market participation* (Discussion Paper No. 21-015). Leibniz Centre for European Economic Research. http://ftp.zew.de/pub/zew-docs/dp/dp21015.pdf

Bucher-Koenen, T., Lusardi, A., Alessie, R. & van Rooij, M. (2017). How financially literate are women? An overview and new insights. *Journal of Consumer Affairs*, *51*(2), 255–283. https://doi.org/10.1111/joca.12121

Buchholz, J. & Hartig, J. (2019). Comparing attitudes across groups: An IRT-based item-fit statistic for the analysis of measurement invariance. *Applied Psychological Measurement*, *43*(3), 241–250. https://doi.org/10.1177/0146621617748323

Cao-Alvira, J. J., Novoa-Hoyos, A. & Núñez-Torres, A. (2020). On the financial literacy, indebtedness, and wealth of Colombian households. *Review of Development Economics*, 1–16. https://doi.org/10.1111/rode.12739

Carlin, B. I. & Robinson, D. T. (2012). What does financial literacy training teach us? *Journal of Economic Education*, *43*(3), 235–247. https://doi.org/10.1080/00220485.2012.686385

Caro, D. H. & Biecek, P. (2017). intsvy: An R package for analyzing international large-scale assessment data. *Journal of Statistical Software*, *81*(7), 1–44. https://doi.org/10.18637/jss.v081.i07

Chambers, R. G. & Asarta, C. J. (2018). Gender, country-level variables, and financial knowledge. *Empirische Pädagogik*, *32*(3/4), 310–328. https://bit.ly/2Pw1HFs

Chambers, R. G., Asarta, C. J. & Farley-Ripple, E. N. (2019). Gender, parental characteristics, and financial knowledge of high school students: Evidence from multicountry data.

Journal of Financial Counseling and Planning, *30*(1), 91–109. https://files.eric.ed.gov/fulltext/EJ1241100.pdf

Chen, Z. & Garand, J. C. (2018). On the gender gap in financial knowledge: Decomposing the effects of don't know and incorrect responses. *Social Science Quarterly*, *99*(5), 1551–1571. https://doi.org/10.1111/ssqu.12520

Chou, C.-P., Bentler, P. M. & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A monte carlo study. *British Journal of Mathematical and Statistical Psychology*, *44*(2), 347–357. https://doi.org/10.1111/j.2044-8317.1991.tb00966.x

Chou, C.-P. & Bentler, P. (1993). Invariant standardized estimated parameter change for model modification in covariance structure analysis. *Multivariate Behavioral Research*, *28*(1), 97–110. https://doi.org/10.1207/s15327906mbr2801_6

Claeskens, G. & Consentino, F. (2008). Variable selection with incomplete covariate data. *Biometrics*, *64*, 1062–1069. https://doi.org/10.1111/j.1541-0420.2008.01003.x

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cole, S., Paulson, A. & Shastry, G. K. (2016). High school curriculum and financial outcomes: The impact of mandated personal finance and mathematics courses. *Journal of Human Resources*, *51*(3), 656–698. https://doi.org/10.3368/jhr.51.3.0113-5410r1

Cole, S., Sampson, T. & Zia, B. (2009). *Financial literacy, financial decisions, and the demand for financial services: Evidence from India and Indonesia* (Working Paper 09-117). Harvard Business School. http://www1.worldbank.org/prem/poverty/ie/dime_papers/1107.pdf

Collins, J. M. (2013). The impacts of mandatory financial education: Evidence from a randomized field study. *Journal of Economic Behavior & Organization*, *95*, 146–158. https://doi.org/10.1016/j.jebo.2012.08.011

Cornell, D. G. & Mayer, M. J. (2010). Why do school order and safety matter? *Educational Researcher*, *39*(1), 7–15. https://doi.org/10.3102/0013189x09357616

Cox, R., Brounen, D. & Neuteboom, P. (2015). Financial literacy, risk aversion and choice of mortgage type by households. *Journal of Real Estate Finance and Economics*, *50*, 74–112. https://doi.org/10.1007/s11146-013-9453-9

Cupak, A., Fessler, P., Silgoner, M. & Ulbrich, E. (2018). *Exploring differences in financial literacy across countries: The role of individual characteristics and institutions* (Working Paper 220). Oesterreichische Nationalbank. https://www.oenb.at/dam/jcr:c6506da3-61d3-4be8-abee-cf19468c13fa/WP220.pdf

Curran, P. J., West, S. G. & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*(1), 16–29. https://doi.org/10.1037/1082-989X.1.1.16

Danes, S. M. & Haberman, H. R. (2007). Teen financial knowledge, self-efficacy, and behavior: A gendered view. *Journal of Financial Counseling and Planning*, *18*(2), 48–60. https://files.eric.ed.gov/fulltext/EJ1104367.pdf

von Davier, M. (2014). Imputing proficiency data under planned missingness in population models. In L. Rutkowski, M. von Davier & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis* (pp. 175–201). CRC Press. https://doi.org/10.1201/b16061-13

Davies, P. (2015). Towards a framework for financial literacy in the context of democracy. *Journal of Curriculum Studies*, *47*(2), 300–316. https://doi.org/10.1080/00220272.2014.934717

Drew, J. M. & Cross, C. (2016). Fraud and its PREY: Conceptualising social engineering tactics and its impact on financial literacy outcomes. In T. Harrison (Ed.), *Financial literacy and the limits of financial decision-making* (pp. 325–340). Springer. https://doi.org/10.1007/978-3-319-30886-9_16

Driva, A., Lührmann, M. & Winter, J. (2016). Gender differences and stereotypes in financial literacy: Off to an early start. *Economics Letters*, *146*, 143–146. https://doi.org/10.1016/j.econlet.2016.07.029

Enders, C. K. & Mansolf, M. (2018). Assessing the fit of structural equation models with multiply imputed data. *Psychological Methods*, *23*(1), 76–93. https://doi.org/10.1037/met0000102

Erner, C., Goedde-Menke, M. & Oberste, M. (2016). Financial literacy of high school students: Evidence from Germany. *Journal of Economic Education*, *47*(2), 95–105. https://doi.org/10.1080/00220485.2016.1146102

Esposito, C. (1999). Learning in urban blight: School climate and its effect on the school performance of urban, minority, low-income children. *School Psychology Review*, *28*(3), 365–377. https://doi.org/10.1080/02796015.1999.12085971

Farinella, J., Bland, J. & Franco, J. (2017). The impact of financial education on financial literacy and spending habits. *International Journal of Business, Accounting, and Finance*, *11*(1), 1–12. http://www.iabpad.com/the-impact-of-financial-education-on-financial-literacy-and-spending-habits/

Fernandes, D., Lynch, J. G. & Netemeyer, R. G. (2014). Financial literacy, financial education, and downstream financial behaviors. *Management Science*, *60*(8), 1861–1883. https://doi.org/10.1287/mnsc.2013.1849

Ferragina, E., Seeleib-Kaiser, M. & Spreckelsen, T. (2015). The four worlds of 'welfare reality'—Social risks and outcomes in Europe. *Social Policy and Society*, *14*(2), 287–307. https://doi.org/10.1017/s1474746414000530

Finlayson, A. (2009). Financialisation, financial literacy and asset-based welfare. *The British Journal of Politics and International Relations*, *11*(3), 400–421. https://doi.org/10.1111/j.1467-856x.2009.00378.x

Finn, J. D. & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis*, *21*(2), 97–109. https://doi.org/10.3102/01623737021002097

Fonseca, R., Mullen, K. J., Zamarro, G. & Zissimopoulos, J. (2012). What explains the gender gap in financial literacy? The role of household decision making. *Journal of Consumer Affairs*, *46*(1), 90–106. https://doi.org/10.1111/j.1745-6606.2011.01221.x

Freiberg, H. J. & Stein, T. A. (1999). Measuring, improving and sustaining healthy learning environments. In H. J. Freiberg (Ed.), *School climate: Measuring, improving and sustaining healthy learning environments* (pp. 11–29). RoutledgeFalmer.

Gale, W. G., Harris, B. H. & Levine, R. (2012). Raising household saving: Does financial education work? *Social Security Bulletin*, *72*(2), 39–48. https://www.ssa.gov/policy/docs/ssb/v72n2/v72n2p39.pdf

Gamble, K. J., Boyle, P. A., Yu, L. & Bennett, D. A. (2015). Aging and financial decision making. *Management Science*, *61*(11), 2603–2610. https://doi.org/10.1287/mnsc.2014.2010

Gee, J. (2018). *Annual fraud indicator 2017: Identifying the cost of fraud to the UK economy.* Growe UK. https://www.crowe.com/uk/croweuk/-/media/Crowe/Firms/Europe/uk/CroweUK/PDF-publications/Annual-Fraud-Indicator-report-2017

Geiger, V., Yasukawa, K., Bennison, A., Wells, J. F. & Sawatzki, C. (2020). Facets of numeracy: Teaching, learning and practices. In J. Way, C. Attard, J. Anderson, J. Bobis, H. McMaster & K. Cartwright (Eds.), *Research in mathematics education in Australasia 2016–2019* (pp. 59–89). Springer. https://doi.org/10.1007/978-981-15-4269-5_4

Gerardi, K., Goette, L. & Meier, S. (2010). Financial literacy and subprime mortgage delinquency: Evidence from a survey matched to administrative data [Working paper 2010-10]. *Federal Reserve Bank of Atlanta Working Paper Series.* https://www.atlantafed.org/-/media/documents/research/publications/wp/2010/wp1010.pdf

Gilbert, N. (2002). *Transformation of the welfare state: The silent surrender of public responsibility.* Oxford University Press.

Gottfredson, G. D., Gottfredson, D. C., Payne, A. A. & Gottfredson, N. C. (2005). School climate predictors of school disorder: Results from a national study of delinquency prevention in schools. *Journal of Research in Crime and Delinquency*, *42*(4), 412–444. https://doi.org/10.1177/0022427804271931

Gramaţki, I. (2017). A comparison of financial literacy between native and immigrant school students. *Education Economics*, *25*(3), 304–322. https://doi.org/10.1080/09645292.2016.1266301

Grohmann, A. (2016). The gender gap in financial literacy: Income, education, and experience offer only partial explanations. *DIW Economic Bulletin*, *6*(46/47), 531–537. https://www.econstor.eu/bitstream/10419/148080/1/872886581.pdf

Grund, S., Lüdtke, O. & Robitzsch, A. (2016). Multiple imputation of multilevel missing data: An introduction to the R package pan. *SAGE Open*, 1–17. https://doi.org/10.1177/2158244016668220

Grund, S., Lüdtke, O. & Robitzsch, A. (2018). Multiple imputation of missing data for multilevel models: Simulations and recommendations. *Organizational Research Methods*, *21*(1), 111–149. https://doi.org/10.1177/1094428117703686

Gudmunson, C. G. & Danes, S. M. (2011). Family financial socialization: Theory and critical review. *Journal of Family and Economic Issues*, *32*(4), 644–667. https://doi.org/10.1007/s10834-011-9275-y

Harlen, W. (1999). Purposes and procedures for assessing science process skills. *Assessment in Education: Principles, Policy & Practice*, *6*(1), 129–144. https://doi.org/10.1080/09695949993044

Hastings, J. S., Madrian, B. C. & Skimmyhorn, W. L. (2013). Financial literacy, financial education, and economic outcomes. *Annual Review of Economics*, *5*(1), 347–373. https://doi.org/10.1146/annurev-economics-082312-125807

Hospido, L., Izquierdo, S. & Machelett, M. (2021). *The gender gap in financial competences* (Economic Bulletin 1/2021). Banco de España. https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/InformesBoletinesRevistas/ArticulosAnaliticos/21/T1/descargar/Files/be2101-art05e.pdf

Hox, J. J., Maas, C. J. M. & Brinkhuis, M. J. S. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica*, *64*(2), 157–170. https://doi.org/10.1111/j.1467-9574.2009.00445.x

Hu, L.-t. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In J. M. Le Cam & J. Neyman (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (pp. 221–233). University of California Press. https://digitalassets.lib.berkeley.edu/math/ucb/text/math_s5_v1_article-13.pdf

Huston, S. J. (2010). Measuring financial literacy. *Journal of Consumer Affairs*, *44*(2), 296–316. https://doi.org/10.1111/j.1745-6606.2010.01170.x

Huston, S. J. (2012). Financial literacy and the cost of borrowing. *International Journal of Consumer Studies*, *36*(5), 566–572. https://doi.org/10.1111/j.1470-6431.2012.01122.x

Indefenso, E. E. & Yazon, A. D. (2020). Numeracy level, mathematics problem skills, and financial literacy. *Universal Journal of Educational Research*, *8*(10), 4393–4399. https://doi.org/10.13189/ujer.2020.081005

Jappelli, T. (2010). Economic literacy: An international comparison. *The Economic Journal*, *120*(548), F429–F451. https://doi.org/10.1111/j.1468-0297.2010.02397.x

Jimerson, S. R., Hart, S. R. & Renshaw, T. L. (2012). Conceptual foundations for understanding youth engaged in antisocial and aggressive behaviors. In S. Jimerson, A. Nickerson,

M. J. Mayer & M. J. Furlong (Eds.), *Handbook of school violence and school safety* (2nd ed., pp. 3–14). Routledge. https://doi.org/10.4324/9780203841372

Jorgensen, B. L. & Savla, J. (2010). Financial literacy of young adults: The importance of parental socialization. *Family Relations*, *59*(4), 465–478. https://doi.org/10.1111/j.1741-3729.2010.00616.x

Kaiser, T. & Menkhoff, L. (2020). Financial education in schools: A meta-analysis of experimental studies. *Economics of Education Review*, *78*, 1–15. https://doi.org/10.1016/j.econedurev.2019.101930

Kaplan, D. (1989). Model modification in covariance structure analysis: Application of the expected parameter change statistic. *Multivariate Behavioral Research*, *24*(3), 285–305. https://doi.org/10.1207/s15327906mbr2403_2

Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions* (2nd ed.). SAGE. https://doi.org/10.4135/9781452226576

Karakurum-Ozdemir, K., Kokkizil, M. & Uysal, G. (2019). Financial literacy in developing countries. *Social Indicators Research*, *143*(1), 325–353. https://doi.org/10.1007/s11205-018-1952-x

Khoirunnisaa, J. & Johan, I. R. (2020). The effects of financial literacy and self-control towards financial behavior among high school students in Bogor. *Journal of Consumer Sciences*, *5*(2), 73–86. https://doi.org/10.29244/jcs.5.2.73-86

Kiliyanni, A. L. & Sivaraman, S. (2016). The perception-reality gap in financial literacy: Evidence from the most literate state in India. *International Review of Economics Education*, *23*, 47–64. https://doi.org/10.1016/j.iree.2016.07.001

Klapper, L. & Lusardi, A. (2019). Financial literacy and financial resilience: Evidence from around the world. *Financial Management*, 1–26. https://doi.org/10.1111/fima.12283

Krippner, G. R. (2005). The financialization of the American economy. *Socio-Economic Review*, *3*(2), 173–208. https://doi.org/10.1093/ser/mwi008

Kuperminc, G. P., Leadbeater, B. J., Emmons, C. & Blatt, S. J. (1997). Perceived school climate and difficulties in the social adjustment of middle school students. *Applied Developmental Science*, *1*(2), 76–88. https://doi.org/10.1207/s1532480xads0102_2

Kutsyuruba, B., Klinger, D. A. & Hussain, A. (2015). Relationships among school climate, school safety, and student achievement and well-being: A review of the literature. *Review of Education*, *3*(2), 103–135. https://doi.org/10.1002/rev3.3043

Lambert, B. (2018). *A student's guide to Bayesian statistics*. SAGE.

Lee, J. & Stankov, L. (2018). Non-cognitive predictors of academic achievement: Evidence from TIMSS and PISA. *Learning and Individual Differences*, *65*, 50–64. https://doi.org/10.1016/j.lindif.2018.05.009

Little, R. J. A. & Rubin, D. B. (2019). *Statistical analysis with missing data* (3rd ed.). Wiley. https://doi.org/10.1002/9781119482260

Longobardi, S., Pagliuca, M. M. & Regoli, A. (2018). Can problem-solving attitudes explain the gender gap in financial literacy? Evidence from Italian students' data. *Quality & Quantity*, *52*(4), 1677–1705. https://doi.org/10.1007/s11135-017-0545-0

Lüdtke, O., Marsh, H. W., Robitzsch, A. & Trautwein, U. (2011). A 2×2 taxonomy of multilevel latent contextual models: Accuracy–bias trade-offs in full and partial error correction models. *Psychological Methods*, *16*(4), 444–467. https://doi.org/10.1037/a0024376

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T. & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*(3), 203–229. https://doi.org/10.1037/a0012869

Lüdtke, O., Robitzsch, A., Trautwein, U. & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology*, *34*, 120–131. https://doi.org/10.1016/j.cedpsych.2008.12.001

Lührmann, M., Serra-Garcia, M. & Winter, J. (2015). Teaching teenagers in finance: Does it work? *Journal of Banking & Finance*, *54*, 160–174. https://doi.org/10.1016/j.jbankfin.2014.11.009

Lusardi, A. (2012). Financial literacy and financial decision-making in older adults. *Generations*, *36*(2), 25–32. https://www.jstor.org/stable/26555907

Lusardi, A. (2015). Financial literacy skills for the 21st Century: Evidence from PISA. *Journal of Consumer Affairs*, *49*(3), 639–659. https://doi.org/10.1111/joca.12099

Lusardi, A. & Mitchell, O. S. (2007). Baby Boomer retirement security: The roles of planning, financial literacy, and housing wealth. *Journal of Monetary Economics*, *54*(1), 205–224. https://doi.org/10.1016/j.jmoneco.2006.12.001

Lusardi, A. & Mitchell, O. S. (2008). Planning and financial literacy: How do women fare? *American Economic Review: Papers & Proceedings*, *98*(2), 413–417. https://doi.org/10.1257/aer.98.2.413

Lusardi, A. & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature*, *52*(1), 5–44. https://doi.org/10.1257/jel.52.1.5

Lusardi, A., Mitchell, O. S. & Curto, V. (2010). Financial literacy among the young. *Journal of Consumer Affairs*, *44*(2), 358–380. https://doi.org/10.1111/j.1745-6606.2010.01173.x

Lusardi, A., Schneider, D. & Tufano, P. (2015). The economic crisis and medical care use: Comparative evidence from five high-income countries. *Social Science Quarterly*, *96*(1), 202–213. https://doi.org/10.1111/ssqu.12076

MacCallum, R. C., Roznowski, M. & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*(3), 490–504. https://doi.org/10.1037/0033-2909.111.3.490

Mandell, L. & Klein, L. S. (2009). The impact of financial literacy education on subsequent financial behavior. *Journal of Financial Counseling and Planning*, *20*(1), 15–24. https://files.eric.ed.gov/fulltext/EJ859556.pdf

Marley-Payne, J., Dituri, P. & Davidson, A. (2021). *Financial education and conceptual understanding: Learning from best practices in mathematics*. Financial Life Cycle Education Corp. https://ficycle.org/wp-content/uploads/2021/04/FINAL-Financial-Education-and-Conceptual-Understanding.pdf

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S. & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist*, *47*(2), 106–124. https://doi.org/10.1080/00461520.2012.670488

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B. & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, *44*(6), 764–802. https://doi.org/10.1080/00273170903333665

Meng, X.-L. & Rubin, D. B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika*, *79*(1), 103–111. https://doi.org/10.1093/biomet/79.1.103

Miles, K. H. & Darling-Hammond, L. (1998). Rethinking the allocation of teaching resources: Some lessons from high-performing schools. *Educational Evaluation and Policy Analysis*, *20*(1), 9–29. https://doi.org/10.3102/01623737020001009

Montoya, D. Y. & Scott, M. L. (2013). The effect of lifestyle-based depletion on teen consumer behavior. *Journal of Public Policy & Marketing*, *32*(1), 82–96. https://doi.org/10.1509/jppm.10.086

Moreno-Herrero, D., Salas-Velasco, M. & Sánchez-Campillo, J. (2018). Factors that influence the level of financial literacy among young people: The role of parental engagement and students' experiences with money matters. *Children and Youth Services Review*, *95*, 334–351. https://doi.org/10.1016/j.childyouth.2018.10.042

Morgan, R. E. (2021). *Financial fraud in the United States, 2017*. U.S. Department of Justice. https://www.bjs.gov/content/pub/pdf/ffus17.pdf

Mountain, T. P., Kim, N., Serido, J. & Shim, S. (2020). Does type of financial learning matter for young adults' objective financial knowledge and financial behaviors? A longitudinal and mediation analysis. *Journal of Family and Economic Issues*, 1–20. https://doi.org/10.1007/s10834-020-09689-6

Muthén, L. K. & Muthén, B. O. (2017). ***Mplus** user's guide*. Muthén & Muthén. https://www.statmodel.com/download/usersguide/Mplus%20user%20guide%20Ver_7_r6_web.pdf

Muthén, L. K. & Muthén, B. O. (1998–2017). **Mplus** (Version 8.5) [Computer software]. Muthén & Muthén. http://www.statmodel.com/

Norvilitis, J. M. & MacLean, M. G. (2010). The role of parents in college students' financial behaviors and attitudes. *Journal of Economic Psychology*, *31*(1), 55–63. https://doi.org/10.1016/j.joep.2009.10.003

OECD. (2009). *PISA data analysis manual* (SPSS 2nd ed.). OECD Publishing. https://doi.org/10.1787/9789264056275-en

OECD. (2019a). PISA 2018 financial literacy framework. *PISA 2018 assessment and analytical framework* (pp. 119–164). OECD Publishing. https://doi.org/10.1787/a1fad77c-en

OECD. (2019b). *PISA 2018 results: What school life means for students' lives*. OECD Publishing. https://doi.org/10.1787/acd78851-en

OECD. (2020a). *Financial literacy data file* [Data set]. OECD Publishing. https://webfs.oecd.org/pisa2018/SPSS_STU_FLT.zip

OECD. (2020b). *PISA 2018 results: Are students smart about money?* OECD Publishing. https://doi.org/10.1787/48ebd1ba-en

OECD. (2020c). *PISA 2018 technical report*. OECD Publisher. https://www.oecd.org/pisa/data/pisa2018technicalreport/

OECD. (2020d). *School questionnaire data file* [Data set]. OECD Publisher. https://webfs.oecd.org/pisa2018/SPSS_SCH_QQQ.zip

Opletalová, A. (2015). Financial education and financial literacy in the Czech education system. *Procedia – Social and Behavioral Sciences*, *171*, 1176–1184. https://doi.org/10.1016/j.sbspro.2015.01.229

Ozkale, A. & Erdogan, E. O. (2020a). An analysis of the interaction between mathematical literacy and financial literacy in PISA. *International Journal of Mathematical Education in Science and Technology*, 1–21. https://doi.org/10.1080/0020739x.2020.1842526

Ozkale, A. & Erdogan, E. O. (2020b). A conceptual model for the interaction of mathematical and financial literacies. *International Journal of Progressive Education*, *16*(5), 288–304. https://doi.org/10.29329/ijpe.2020.277.18

Pak, T.-Y. (2018). Financial literacy and high-cost borrowing: Exploring the mechanism. *International Journal of Consumer Studies*, *42*(3), 283–294. https://doi.org/10.1111/ijcs.12429

Pokropek, A. (2016). Introduction to instrumental variables and their application to large-scale assessment data. *Large-scale Assessments in Education*, *4*(4), 1–20. https://doi.org/10.1186/s40536-016-0018-2

Potrich, A. C. G., Vieira, K. M., Coronel, D. A. & Bender Filho, R. (2015). Financial literacy in Southern Brazil: Modeling and invariance between genders. *Journal of Behavioral and Experimental Finance*, *6*, 1–12. https://doi.org/10.1016/j.jbef.2015.03.002

Potrich, A. C. G., Vieira, K. M. & Kirch, G. (2015). Determinants of financial literacy: Analysis of the influence of socioeconomic and demographic variables. *Revista Contabilidade & Finanças*, *26*(69), 362–377. https://doi.org/10.1590/1808-057x201501040

Preston, A. C. & Wright, R. E. (2019). Understanding the gender gap in financial literacy: Evidence from Australia. *Economic Record*, *95*(S1), 1–29. https://doi.org/10.1111/1475-4932.12472

R Core Team. (2021). *R: A language and environment for statistical computing* (Version 4.0.5) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Raykov, T. & Marcoulides, G. A. (2004). Using the delta method for approximate interval estimation of parameter functions in SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(4), 621–637. https://doi.org/10.1207/s15328007sem1104_7

Regan, S. & Paxton, W. (2003). *Beyond bank accounts: Full financial inclusion.* IPPR. https://www.ippr.org/files/images/media/files/publication/2011/05/beyond_bank_accounts_1297.pdf

Reiter, S. & Beckmann, E. (2020). *How financially literate is CESEE? Insights from the OeNB Euro Survey* (tech. rep.). Oesterreichische Nationalbank. https://www.oenb.at/dam/jcr:578c0407-1d22-4094-a312-b7ce3e82ae76/03_feei_Q3_20_How-financially-literate-is-CESEE.pdf

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* John Wiley & Sons. https://doi.org/10.1002/9780470316696

Rust, K. (2014). Samping, weighting, and variance estimation in international large-scale assessments. In L. Rutkowski, M. von Davier & D. Rutkowski (Eds.), *Handbook of international large-sclae assessment: Background, technical issues, and methods of data analysis* (pp. 117–153). CRC Press. https://doi.org/10.1201/b16061-11

Salas-Velasco, M. (2019). The educational performance of Spanish secondary schools in PISA, 2003–2012. *Current Politics and Economics of Europe*, *30*(2), 189–218. https://www.proquest.com/docview/2279748994/fulltext/3AFA8983FDCC4D84PQ/

Saris, W. E., Satorra, A. & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. In C. C. Clogg (Ed.), *Sociological methodology* (pp. 105–129). American Sociological Association. https://doi.org/10.2307/271030

Schuhen, M. & Schürkmann, S. (2014). Construct validity of financial literacy. *International Review of Economics Education*, *16*, 1–11. https://doi.org/10.1016/j.iree.2014.07.004

Seeleib-Kaiser, M. (2016). The end of the conservative German welfare state model. *Social Policy & Administration*, *50*(2), 219–240. https://doi.org/10.1111/spol.12212

Serido, J. & Deenanath, V. (2016). Financial parenting: Promoting financial self-reliance of young consumers. In J. J. Xiao (Ed.), *Handbook of consumer finance research* (pp. 291–300). Springer. https://doi.org/10.1007/978-3-319-28887-1_24

Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth Cengage Learning.

Silgoner, M., Greimel-Fuhrmann, B. & Weber, R. (2015). Financial literacy gaps of the Austrian population. *Monetary Policy & the Economy*, *Q2*, 35–51. https://www.oenb.at/dam/jcr:a23bbdba-3696-4ed8-a4d5-656bbf09e0e0/mop_2015_q2_analyses02.pdf

Skagerlund, K., Lind, T., Strömbäck, C., Tinghög, G. & Västfjäll, D. (2018). Financial literacy and the role of numeracy–How individuals' attitude and affinity with numbers influence financial literacy. *Journal of Behavioral and Experimental Economics*, *74*, 18–25. https://doi.org/10.1016/j.socec.2018.03.004

Skiba, R., Ritter, S., Simmons, A., Peterson, R. & Miller, C. (2006). The safe and responsive schools project: A school reform model for implementing best practices in violence prevention. In S. R. Jimerson & M. Furlong (Eds.), *Handbook of school violence and school safety: From research to practice* (pp. 631–650). Lawrence Erlbaum. https://psycnet.apa.org/record/2006-03632-041

Smith, G., Wood, L., Coupland, M., Stephenson, B., Crawford, K. & Ball, G. (1996). Constructing mathematical examinations to assess a range of knowledge and skills. *International Journal of Mathematical Education in Science and Technology*, *27*(1), 65–77. https://doi.org/10.1080/0020739960270109

Snijders, T. A. B. & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. SAGE.

Sole, M. A. (2014). Financial literacy: An essential component of mathematics literacy and numeracy. *Journal of Mathematics Education at Teachers College*, *5*(2), 55–62. https://academicworks.cuny.edu/cgi/viewcontent.cgi?article=1011&context=nc_pubs

Stone, D. M., Simon, T. R., Fowler, K. A., Kegler, S. R., Yuan, K., Holland, K. M., Ivey-Stephenson, A. Z. & Crosby, A. E. (2018). *Vital signs*: Trends in state suicide rates — United States, 1999–2016 and circumstances contributing to suicide — 27 states, 2015. *Morbidity and Mortality Weekly Report*, *67*(22), 617–624. https://doi.org/10.15585/mmwr.mm6722a1

Sutter, M., Weyand, M., Untertrifaller, A. & Froitzheim, M. (2020). *Financial literacy, risk and time preferences–Results from a randomized educational intervention* (Discussion Paper 2020/17). Max Planck Institute for Research on Collective Goods. https://homepage.coll.mpg.de/pdf_dat/2020_17online.pdf

Tang, N. (2017). Like father like son: How does parents' financial behavior affect their children's financial behavior? *Journal of Consumer Affairs*, *51*(2), 284–311. https://doi.org/10.1111/joca.12122

Taylor, S. M. & Wagland, S. (2013). The solution to the financial literacy problem: What is the answer? *Australasian Accounting, Business and Finance Journal*, *7*(3), 69–90. https://doi.org/10.14453/aabfj.v7i3.5

Teig, N., Scherer, R. & Nilsen, T. (2018). More isn't always better: The curvilinear relationship between inquiry-based teaching and student achievement in science. *Learning and Instruction*, *56*, 20–29. https://doi.org/10.1016/j.learninstruc.2018.02.006

Thomson, S. & De Bortoli, L. (2017). *PISA 2015: Financial literacy in Australia*. Australian Council for Educational Research. https://research.acer.edu.au/cgi/viewcontent.cgi?article=1028&context=ozpisa

Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen & K. Elliot (Eds.), *But what does it* mean*? The use of effect sizes in educational research* (pp. 55–66). NFER; Institute of Education University of London. https://dro.dur.ac.uk/23722/1/23722.pdf

Uline, C. & Tschannen-Moran, M. (2008). The walls speak: The interplay of quality facilities, school climate, and student achievement. *Journal of Educational Administration*, *46*(1), 55–73. https://doi.org/10.1108/09578230810849817

Utkarsh, B., Pandey, A., Ashta, A., Spiegelman, E. & Sutan, A. (2020). Catch them young: Impact of financial socialization, financial literacy and attitude towards money on financial well-being of young adults. *International Journal of Consumer Studies*, *44*(6), 531–541. https://doi.org/10.1111/ijcs.12583

Wang, M.-T. & Degol, J. L. (2016). School climate: A review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review*, *28*(2), 315–352. https://doi.org/10.1007/s10648-015-9319-1

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, *50*(1), 1–25. https://doi.org/10.2307/1912526

Whittaker, T. A. (2012). Using the modification index and standardized expected parameter change for model modification. *Journal of Experimental Education*, *80*(1), 26–44. https://doi.org/10.1080/00220973.2010.531299

Yoshino, N., Morgan, P. J. & Wignaraja, G. (2015). *Financial education in Asia: Assessment and recommendations* (Working Paper Series No. 534). Asian Development Bank Institute. https://www.adb.org/sites/default/files/publication/161053/adbi-wp534.pdf

Zhang, X., Wan, Q., Lyu, S., Li, O. & Liu, Y. (2021). Overlearning is as ineffective as underlearning? A cross-culture study from PISA 2015. *Learning and Individual Differences*, *88*, 1–10. https://doi.org/10.1016/j.lindif.2021.102005

Appendices

Appendix A GDPR Documentation and Ethical Approval

This research project discharges its duty imposed by the EEA's general data protection regulation (GDPR) by following Norwegian Centre for Research Data (NSD)'s notification test on Friday, 11 September 2020. Both PISA 2018 Database and the World Bank Open Data contain only aggregated and de-personalised datasets with no possibility of back-tracing to any particular participant. Resultantly, no identifiable personal data were collected or used at any stage of this research. The NSD's assessment letter outlines the agency's decision of not subjecting this project to the GDPR notification. The NSD decision letter also satisfies University of Oslo's ethical approval requirement and concludes the approval process.

About us (/personvernombud/en/about_us.html)
Norwegian (/personvernombud/meld_prosjekt/meldeplikttest.html)

NSD (/) > Personverntjenester (/personvernombud/) > Data Protection Services (/personvernombud/en/) > Notify project (/personvernombud/en/notify/) > Notification Test

Denne siden på norsk (/personvernombud/meld_prosjekt/meldeplikttest.html)

Will you be processing personal data?

Are you unsure whether your project is subject to notification? Feel free to try our informal Notification test. Note that the test is intended as a guidance and is not a formal assessment.

Will you be collecting/processing directly identifiable personal data? ○ Yes ◉ No

A person will be directly identifiable through name, social security number, or other uniquely personal characteristics.

Read more about personal data (/personvernombud/en/help/vocabulary.html?id=8) and notification (/personvernombud/en/notify/index.html).

NB! Even though information is to be anonymized in the final thesis/report, check the box if identifying personal data is to be collected/processed in connection with the project.

Will directly identifiable personal information be linked to the data (e.g. through a reference number which refers to a separate list of names/scrambling key)? ○ Yes ◉ No

Note that the project will be subject to notification even if you cannot access the scrambling key (/personvernombud/en/help/vocabulary.html?id=11), as the procedure often is when using a data processor (/personvernombud/en/help/vocabulary.html?id=3), or in register-based studies (/personvernombud/en/help/research_methods/register_studies.html).

Will you be collecting/processing background information that may identify individuals (indirectly identifiable personal data)? ○ Yes ◉ No

A person will be indirectly identifiable if it is possible to identify him/her through a combination of background information (such as place of residence or workplace/school, combined with information such as age, gender, occupation, etc.).

Will there be registered personal data (directly/indirectly/via IP or email address, etc.) using online surveys? ○ Yes ◉ No

Please note that the project will be subject to notification even if you as a student/researcher cannot access the link to the IP or email address, as the procedure often is when using a data processor.

Read more about online surveys (/personvernombud/en/help/research_methods/online_surveys.html).

Will there be registered personal data using digital photo or video files? ○ Yes ● No

Photo/video recordings of faces will be regarded as identifiable personal data. In order for a voice to be considered as identifiable, it must be registered in combination with other background information, in such a way that people can be recognized.

Show results

Notify project

Do I have to notify my project? (/personvernombud/en/notify/index.html)

Notification Form (/personvernombud/en/notify/meldeskjema_link)

Notifying changes (/personvernombud/en/notify/notifying_changes.html)

Get help notifying your project

Processing the notification (/personvernombud/en/help/index.html)

Frequently asked questions (/personvernombud/en/help/faq.html)

Vocabulary (/personvernombud/en/help/vocabulary.html)

Research topics (/personvernombud/en/help/research_topics/)

Research methods (/personvernombud/en/help/research_methods/)

Information and consent (/personvernombud/en/help/information_consent/)

Other approvals (/personvernombud/en/help/other_approvals/)

Result of Notification Test: Not Subject to Notification

You have indicated that neither directly or indirectly identifiable personal data will be registered in the project.

If no personal data is to be registered, the project will not be subject to notification, and you will not have to submit a notification form.

Please note that this is a guidance based on information that you have given in the notification test and not a formal confirmation.

For your information: *In order for a project not to be subject to notification, we presuppose that all information processed using electronic equipment in the project remains anonymous.*

Anonymous information is defined as information that cannot identify individuals in the data set in any of the following ways:

- *directly, through uniquely identifiable characteristic (such as name, social security number, email address, etc.)*
- *indirectly, through a combination of background variables (such as residence/institution, gender, age, etc.)*
- *through a list of names referring to an encryption formula or code, or*
- *through recognizable faces on photographs or video recordings.*

Furthermore, we presuppose that names/consent forms are not linked to sensitive personal data.

Kind regards,
NSD Data Protection

Appendix B Analysis Code, Additional Tables and Figures

Full analysis code can be obtained from the author's GitHub page:

https://github.com/tonyctan/CEMO-master-thesis

B.1 Data Merging

```r
# Import SPSS file into R
library(intsvy)
finlit <- pisa.select.merge(
    student.file = "CY07_MSU_FLT_QQQ.SAV", # file ext in capital
    school.file = "CY07_MSU_SCH_QQQ.sav", # file ext in lower case
    student = c(
    # Control variables
        "ST004D01T", # Student (Standardized) Gender
        "IMMIG", # Index Immigration status
        "ESCS", # Index of economic, social and cultural status
    # Mediators
        "FCFMLRTY", # Familiarity with concepts of finance (Sum)
        "FLCONFIN", # Confidence about financial matters (WLE)
    # Academic
        "FLSCHOOL", # Financial education in school lessons (WLE)
    # Safety
        "BEINGBULLIED", # Student's experience of being bullied (WLE)
    # Community
        "FLFAMILY" # Parental involvement in matters of Financial Literacy (WLE)
    ),
    school = c(
        "STRATIO", # Student-teacher ratio
        "EDUSHORT" # Shortage of educational material (WLE)
    ),
    countries = c(
        "BRA", "BGR", "CAN", "CHL", "EST",
        "FIN", "GEO", "IDN", "ITA", "LVA",
        "LTU", "NLD", "PER", "POL", "PRT",
        "RUS", "QMR", "QRT", # Russian Federation and other regions
        "SRB", "SVK", "ESP", "USA"
    )
)

names(finlit)
# Throw away columns that I do not need
finlit <- finlit[, -c(5, 7:86)] # 5 = BOOKID; 7:86 = resampling weights
names(finlit)

# Some var need recording
library(car)

# Re-code Russian territories to RUS
finlit$CNT <- recode(finlit$CNT, "
    'QMR' = 'RUS';
    'QRT' = 'RUS'
")

finlit$CNTRYID <- recode(finlit$CNTRYID, "
    982 = 643;
    983 = 643
")
```

```r
# Input country-level FKI
FKI ← recode(finlit$CNT, "
    'NLD' = 0.940;
    'USA' = 0.937;
    'CAN' = 0.784;
    'ITA' = 0.762;
    'FIN' = 0.724;
    'ESP' = 0.627;
    'LTU' = 0.613;
    'PRT' = 0.591;
    'BGR' = 0.583;
    'EST' = 0.577;
    'SVK' = 0.559;
    'POL' = 0.555;
    'LVA' = 0.550;
    'CHL' = 0.544;
    'RUS' = 0.450;
    'GEO' = 0.424;
    'SRB' = 0.423;
    'PER' = 0.309;
    'BRA' = 0.141;
    'IDN' = 0.122
")

# Recode ST004D01T from Sex to Male
MALE ← finlit$ST004D01T - 1

# Revert coding direction: bigger number => safer school
NOBULLY ← finlit$BEINGBULLIED * (-1)

# Recode IMMIG to 1st and 2nd generation
IMMI1GEN ← recode(finlit$IMMIG, "
    1 = 0;
    2 = 0;
    3 = 1
")

IMMI2GEN ← recode(finlit$IMMIG, "
    1 = 0;
    2 = 1;
    3 = 0
")

# Stitch spreadsheet together
names(finlit)
finlit ← cbind(
    FKI, finlit[, c(2:35)], MALE, IMMI1GEN, IMMI2GEN,
    finlit[, c(38:41)], NOBULLY, finlit[, c(43:46)]
)
head(finlit)
names(finlit)

# Remove cases whose school weights (col #45) are NA
obs0 ← dim(finlit)[1]
finlit ← finlit[complete.cases(finlit$W_FSTUWT_SCH_SUM), ]
obs1 ← dim(finlit)[1]
obs0 - obs1 # 12 cases contained missing school weights and have been dropped
rm(obs0, obs1)
```

Table B.1

Summary of Participating Countries

Country	Country	Country	School		Student		Male	
ID	code	name	n	%	n	%	n	%
76	BRA	Brazil	595	8.97	8,310	7.75	4,045	48.68
100	BGR	Bulgaria	197	2.97	4,110	3.84	2,147	52.24
124	CAN	Canada	492	7.42	7,762	7.24	3,858	49.70
152	CHL	Chile	251	3.79	4,482	4.18	2,254	50.29
233	EST	Estonia	229	3.45	4,166	3.89	2,080	49.93
246	FIN	Finland	204	3.08	4,328	4.04	2,199	50.81
268	GEO	Georgia	319	4.81	4,320	4.03	2,239	51.83
360	IND	Indonesia	395	5.96	7,132	6.66	3,454	48.43
380	ITA	Italy	539	8.13	9,182	8.57	4,706	51.25
428	LVA	Latvia	307	4.63	3,151	2.94	1,587	50.36
440	LTU	Lithuania	349	5.26	4,075	3.80	2,060	50.55
528	NLD	The Netherlands	151	2.28	3,042	2.84	1,549	50.92
604	PER	Peru	337	5.08	4,732	4.42	2,390	50.51
616	POL	Poland	235	3.54	4,294	4.01	2,080	48.44
620	PRT	Portugal	276	4.16	4,568	4.26	2,320	50.79
643	RUS	Russian Federation	558	8.42	9,124	8.51	4,601	50.43
688	SRB	Serbia	186	2.81	3,874	3.62	1,951	50.36
703	SVK	Slovak Republic	357	5.38	3,411	3.18	1,683	49.34
724	ESP	Spain	491	7.40	9,361	8.74	4,695	50.15
840	USA	The USA	163	2.46	3,738	3.49	1,871	50.05
		Total	6,631	100	107,162	100	53,769	50.18

χ^2 goodness-of-fit test	School		Student		Male	
	χ^2_{19}	p	χ^2_{19}	p	χ^2_{19}	p
	1,105.8	$< .001$	16,984	$< .001$	20.9	.34

Note. Twelve observations with missing school weights were removed. χ^2 goodness-of-fit tests revealed that the data set was balanced in sex, but not all countries contributed equally to school and student counts.

Table B.2

Scale Reliabilities (Cronbach's alphas) and Item Parameter References for Derived Variables based on IRT Scaling

Country ID	Country code	Country name	School climate variable FLSCHOOL	FLFAMILY	NOBULLY	EDUSHORT	Financial literacy FLCONFIN
76	BRA	Brazil	.896	.871	.794	.858	.929
100	BGR	Bulgaria	.912	.836	.851	.814	.927
124	CAN	Canada	.904	.856	.758	.816	.900
152	CHL	Chile	.885	.851	.784	.818	.915
233	EST	Estonia	.865	.833	.709	.752	.872
246	FIN	Finland	.883	.819	.760	.783	.896
268	GEO	Georgia	.891	.834	.846	.862	.920
360	IND	Indonesia	.878	.827	.756	.892	.931
380	ITA	Italy	.857	.798	.795	.840	.898
428	LVA	Latvia	.846	.813	.703	.780	.897
440	LTU	Lithuania	.909	.869	.846	.779	.921
528	NLD	The Netherlands	.849	.792	.638	.792	.874
604	PER	Peru	.847	.813	.758	.882	.903
616	POL	Poland	.878	.830	.771	.839	.913
620	PRT	Portugal	.896	.844	.775	.849	.899
643	RUS	Russian Federation	.892	.855	.726	.874	.911
688	SRB	Serbia	.926	.853	.838	.786	.939
703	SVK	Slovak Republic	.874	.829	.783	.799	.907
724	ESP	Spain	.879	.812	.779	.854	.912
840	USA	The USA	.908	.839	.756	.881	.909
Reference for scale reliabilities [a]		OECD countries	16.89	16.89	16.58	16.63	16.89
		Partner countries	16.90	16.90	16.59	16.64	16.90
Reference for item parameters [b]			16.93	16.94	16.61	16.66	16.91

Note. [a] [b] Worksheet names in the associated Excel file accompanying Chapter 16 of *PISA 2018 Technical Report* (OECD, 2020c).

B.2 Multilevel Multiple Imputation

B.2.1 *Mplus* Input

```
TITLE:
    Multilevel multiple imputation using JM-AM H1   ! Unrestricted var-cov


DATA:
    file = "~/finlit.dat";


VARIABLE:
    names =
        FKI CNTRYID CNTSCHID CNTSTUID W_STU         ! Administrative vars
        PV1MATH PV2MATH PV3MATH PV4MATH PV5MATH     ! Plausible values for MATH
        PV6MATH PV7MATH PV8MATH PV9MATH PV10MATH
        PV1READ PV2READ PV3READ PV4READ PV5READ     ! Plausible values for READ
        PV6READ PV7READ PV8READ PV9READ PV10READ
        PV1FLIT PV2FLIT PV3FLIT PV4FLIT PV5FLIT     ! Plausible values for FLIT
        PV6FLIT PV7FLIT PV8FLIT PV9FLIT PV10FLIT
        MALE IMMI1GEN IMMI2GEN ESCS                 ! Demographic info
        FCFMLRTY FLCONFIN                           ! Affects
        FLSCHOOL                                    !   Lat var "Academic"
        NOBULLY                                     !   Lat var "Safety"
        FLFAMILY                                    !   Lat var "Community"
        W_SCH STRATIO                               ! School characteristics
        EDUSHORT                                    !   Lat var "inst. env."
        ;

    usevar =                                        ! Var to be imputed
        MALE IMMI1GEN IMMI2GEN ESCS
        FCFMLRTY FLCONFIN
        FLSCHOOL NOBULLY FLFAMILY
        STRATIO EDUSHORT
        ;

    within =                                        ! Amongst which, L1 var are
        MALE IMMI1GEN IMMI2GEN ESCS
        FCFMLRTY FLCONFIN
        FLSCHOOL NOBULLY FLFAMILY
        ;

    between =                                       ! L2 are
        STRATIO EDUSHORT
        ;

    auxiliary =                                     ! Var not participating in
        PV1MATH PV2MATH PV3MATH PV4MATH PV5MATH     ! MI but still to be
        PV6MATH PV7MATH PV8MATH PV9MATH PV10MATH    ! included in final output
        PV1READ PV2READ PV3READ PV4READ PV5READ
        PV6READ PV7READ PV8READ PV9READ PV10READ    ! PVs are already "guesses"
        PV1FLIT PV2FLIT PV3FLIT PV4FLIT PV5FLIT     ! themselves so do NOT use
        PV6FLIT PV7FLIT PV8FLIT PV9FLIT PV10FLIT    ! PVs to guess others
        FKI CNTRYID CNTSTUID W_STU
        W_SCH                                       ! Admin vars
        ;

    cluster = CNTSCHID;

    missing = all (-99);


ANALYSIS:
    processors = 64;                                ! Use all cores of HPC

    type = twolevel;
    estimator = Bayes;

    fbiterations = 50000;                           ! Number of burn-in
```

```
    chains = 4;                                 ! Verify convergence
    bseed = 1234;                               ! For replication study

DATA IMPUTATION:
    impute =
        MALE (c) IMMI1GEN (c) IMMI2GEN (c) ESCS     ! Categoricals have (c)
        FCFMLRTY FLCONFIN
        FLSCHOOL NOBULLY FLFAMILY
        STRATIO EDUSHORT
        ;

    ndatasets = 10;                             ! To merge with 10 PVs
    save = FLIT_MMI_*.dat;
    thin = 5000;                                ! To Avoid autocorrelation

SAVEDATA:
    bpar = bpar.dat;                            ! Capture Bayesian paths

PLOT:
    type = plot2;                               ! For R's MplusAutomation
```

B.2.2 Selected M*plus* Output

```
MODEL FIT INFORMATION

Number of Free Parameters                              22

Bayesian Posterior Predictive Checking using Chi-Square

          95% Confidence Interval for the Difference Between
          the Observed and the Replicated Chi-Square Values

                              28408.938         28906.315

          Posterior Predictive P-Value              0.000

Information Criteria

          Deviance (DIC)                      2100842.641
          Estimated Number of Parameters (pD)      22.054



MODEL RESULTS

                                Posterior  One-Tailed         95% C.I.
                    Estimate       S.D.      P-Value   Lower 2.5%  Upper 2.5%  Significance

Within Level

 Means
    MALE               0.502       0.002      0.000       0.499       0.505      *
    IMMI1GEN           0.029       0.001      0.000       0.028       0.030      *
    IMMI2GEN           0.042       0.001      0.000       0.041       0.044      *
    ESCS              -0.241       0.003      0.000      -0.247      -0.234      *
    FCFMLRTY           7.049       0.017      0.000       7.015       7.083      *
    FLCONFIN          -0.072       0.003      0.000      -0.079      -0.065      *
    FLSCHOOL           0.018       0.003      0.000       0.011       0.024      *
    NOBULLY           -0.059       0.004      0.000      -0.067      -0.052      *
    FLFAMILY           0.064       0.003      0.000       0.057       0.070      *

 Variances
    MALE               0.250       0.001      0.000       0.248       0.252      *
    IMMI1GEN           0.028       0.000      0.000       0.028       0.028      *
    IMMI2GEN           0.041       0.000      0.000       0.040       0.041      *
    ESCS               1.183       0.005      0.000       1.173       1.193      *
    FCFMLRTY          29.753       0.134      0.000      29.494      30.016      *
    FLCONFIN           1.034       0.005      0.000       1.025       1.044      *
    FLSCHOOL           1.040       0.005      0.000       1.031       1.049      *
    NOBULLY            1.110       0.005      0.000       1.100       1.121      *
```

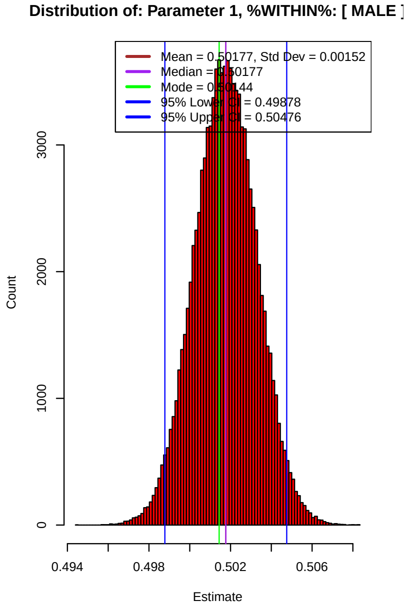
```
    FLFAMILY           1.090        0.005       0.000        1.080        1.100       *

Between Level

 Means
    STRATIO           13.873        0.136       0.000       13.608       14.140       *
    EDUSHORT           0.131        0.013       0.000        0.106        0.157       *

 Variances
    STRATIO          103.514        1.948       0.000       99.805      107.425       *
    EDUSHORT           1.074        0.019       0.000        1.038        1.112       *
```

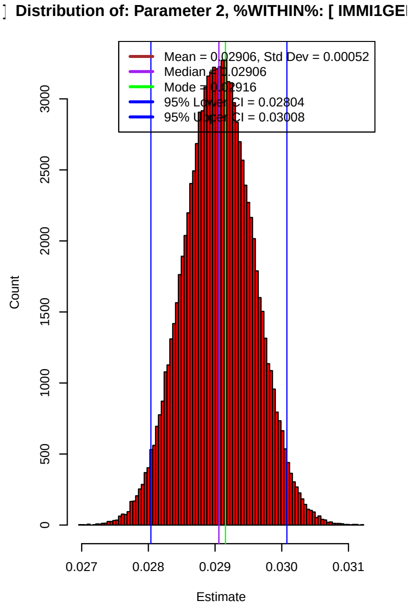
Table B.3

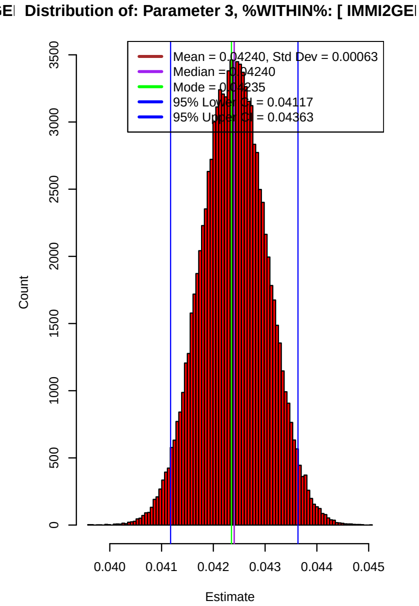
Summary of Diagnostic Plots of Multilevel Multiple Imputation

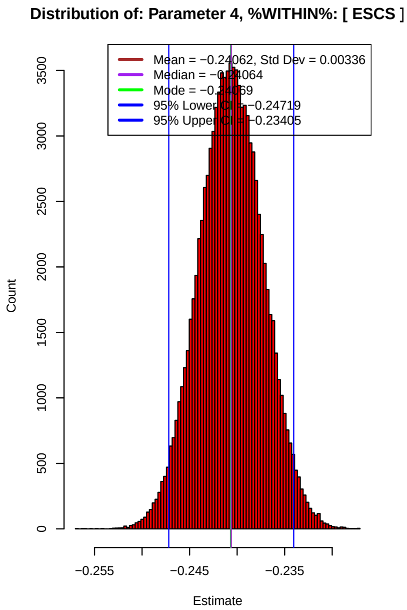
Parameter number	Parameter label	Modelling level	Brief description	Posterior mean	Posterior variance	95% credibility interval	Chain converged	AR-free chains
1	MALE	Within	Whether participant is male	0.502		(0.499, 0.505)	Yes	4
2	IMMI1GEN	Within	Whether participant migrated to this country	0.029		(0.028, 0.030)	Yes	4
3	IMMI2GEN	Within	Whether their parent did	0.042		(0.041, 0.044)	Yes	4
4	ESCS	Within	Index of economic, social and cultural status	−0.241		(−0.247, −0.234)	Yes	4
5	FCFMLRTY	Within	Familiarity with concepts of finance	7.049		(7.015, 7.083)	Yes	4
6	FLCONFIN	Within	Confidence about financial matters	−0.072		(−0.079, −0.065)	Yes	4
7	FLSCHOOL	Within	Financial education in school lessons	0.018		(0.011, 0.024)	Yes	4
8	NOBULLY	Within	Participant's experience of being bullied (reverse)	−0.059		(−0.067, −0.052)	Yes	4
9	FLFAMILY	Within	Parental involvement in matters of financial literacy	0.064		(0.057, 0.070)	Yes	4
10	MALE	Within	Whether participant is male		0.250	(0.248, 0.252)	Yes	4
11	IMMI1GEN	Within	Whether participant migrated to this country		0.028	(0.028, 0.028)	Yes	4
12	IMMI2GEN	Within	Whether their parent		0.041	(0.040, 0.041)	Yes	4
13	ESCS	Within	Index of economic, social and cultural status		1.183	(1.173, 1.193)	Yes	4
14	FCFMLRTY	Within	Familiarity with concepts of finance		29.754	(29.495, 30.016)	Yes	4
15	FLCONFIN	Within	Confidence about financial matters		1.034	(1.025, 1.044)	Yes	4
16	FLSCHOOL	Within	Financial education in school lessons		1.040	(1.031, 1.049)	Yes	4
17	NOBULLY	Within	Participant's experience of being bullied (reverse)		1.111	(1.100, 1.121)	Yes	4
18	FLFAMILY	Within	Parental involvement in matters of financial literacy		1.090	(1.080, 1.100)	Yes	4
19	STRAIO	Between	Student−teacher ratio	13.873		(13.607, 14.139)	Yes	4
20	EDUSHORT	Between	Shortage of educational material	0.131		(0.106, 0.157)	Yes	4
21	STRAIO	Between	Student−teacher ratio		103.532	(99.750, 107.430)	Yes	4
22	EDUSHORT	Between	Shortage of educational material		1.074	(1.037, 1.112)	Yes	4

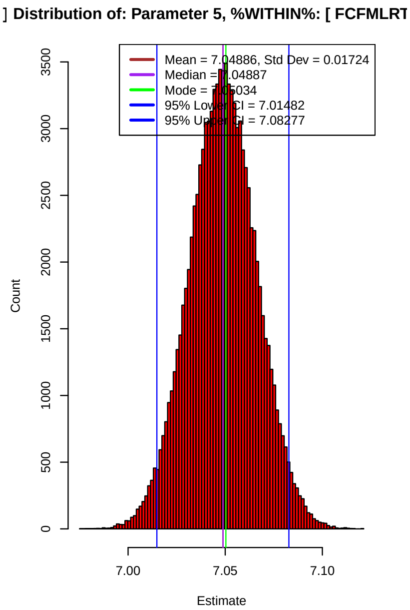
Note. Notes go here.

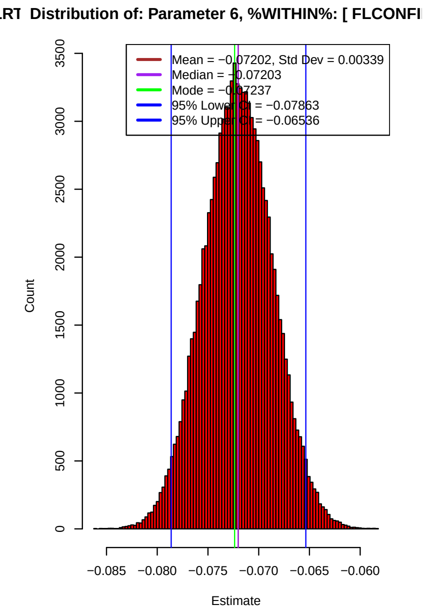

Distribution of: Parameter 1, %WITHIN%: [MALE]
Mean = 0.50177, Std Dev = 0.00152
Median = 0.50177
Mode = 0.50144
95% Lower CI = 0.49878
95% Upper CI = 0.50476
Count
Estimate

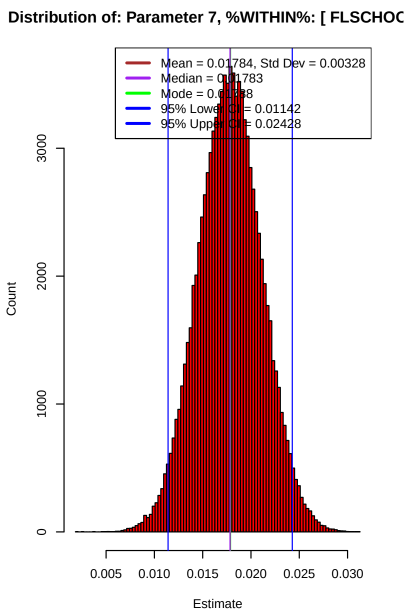

Distribution of: Parameter 2, %WITHIN%: [IMMI1GEI
Mean = 0.02906, Std Dev = 0.00052
Median = 0.02906
Mode = 0.02916
95% Lower CI = 0.02804
95% Upper CI = 0.03008
Count
Estimate

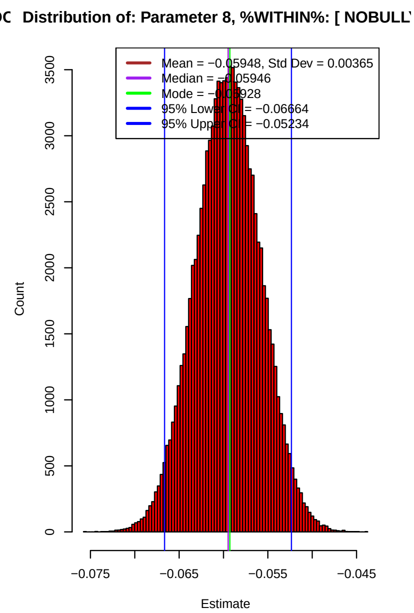

Distribution of: Parameter 3, %WITHIN%: [IMMI2GEI
Mean = 0.04240, Std Dev = 0.00063
Median = 0.04240
Mode = 0.04235
95% Lower CI = 0.04117
95% Upper CI = 0.04363
Count
Estimate

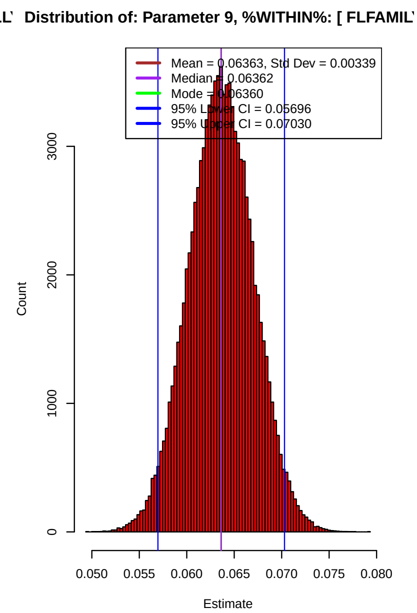

Distribution of: Parameter 4, %WITHIN%: [ESCS]
Mean = −0.24062, Std Dev = 0.00336
Median = −0.24064
Mode = −0.24059
95% Lower CI = −0.24719
95% Upper CI = −0.23405
Count
Estimate

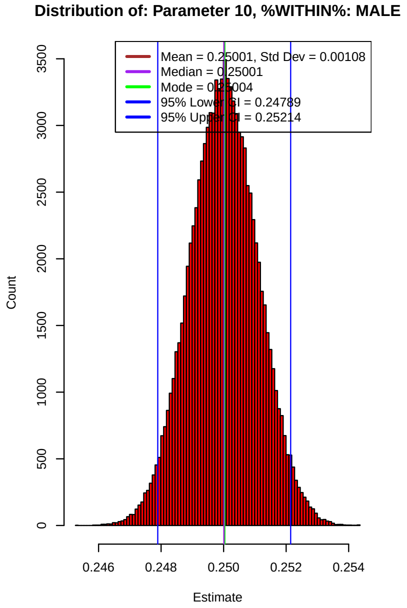

Distribution of: Parameter 5, %WITHIN%: [FCFMLRT
Mean = 7.04886, Std Dev = 0.01724
Median = 7.04887
Mode = 7.05034
95% Lower CI = 7.01482
95% Upper CI = 7.08277
Count
Estimate

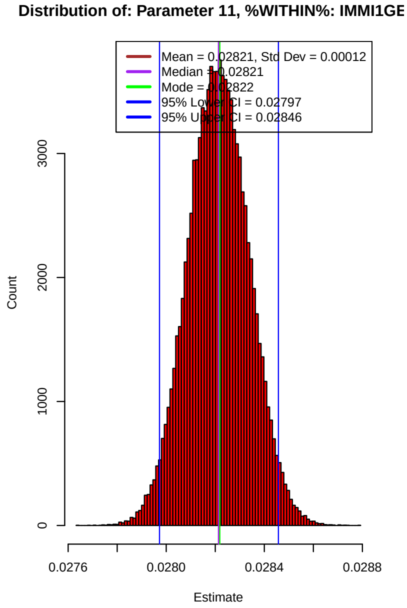

Distribution of: Parameter 6, %WITHIN%: [FLCONFI
Mean = −0.07202, Std Dev = 0.00339
Median = −0.07203
Mode = −0.07237
95% Lower CI = −0.07863
95% Upper CI = −0.06536
Count
Estimate

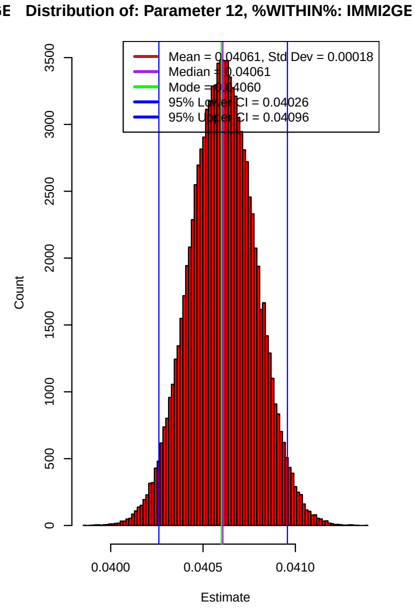

Distribution of: Parameter 7, %WITHIN%: [FLSCHOC
Mean = 0.01784, Std Dev = 0.00328
Median = 0.01783
Mode = 0.01788
95% Lower CI = 0.01142
95% Upper CI = 0.02428
Count
Estimate

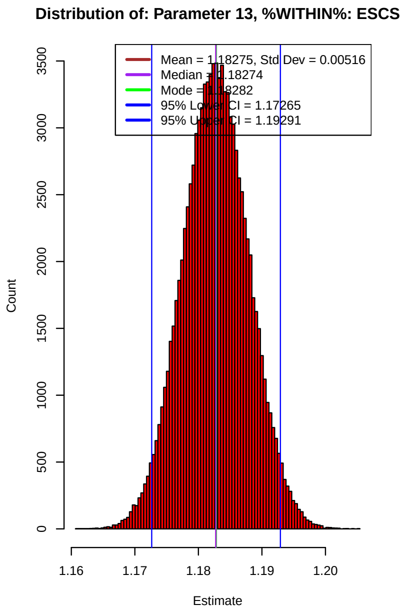

Distribution of: Parameter 8, %WITHIN%: [NOBULL
Mean = −0.05948, Std Dev = 0.00365
Median = −0.05946
Mode = −0.05928
95% Lower CI = −0.06664
95% Upper CI = −0.05234
Count
Estimate

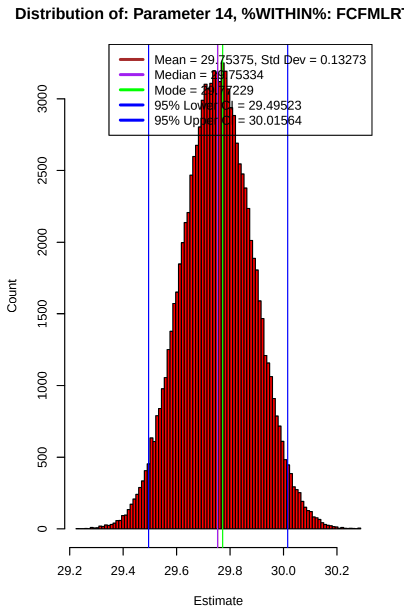

Distribution of: Parameter 9, %WITHIN%: [FLFAMIL
Mean = 0.06363, Std Dev = 0.00339
Median = 0.06362
Mode = 0.06360
95% Lower CI = 0.05696
95% Upper CI = 0.07030
Count
Estimate

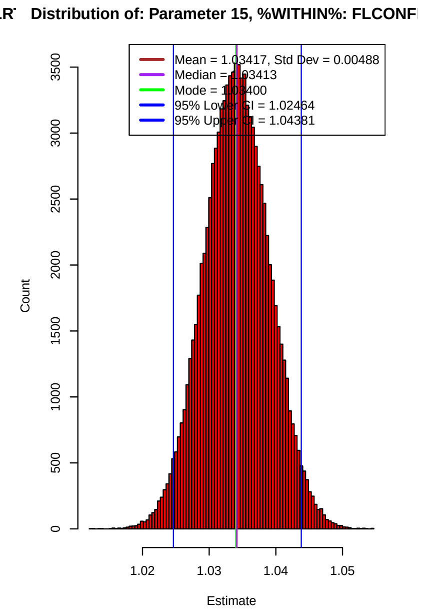

Distribution of: Parameter 10, %WITHIN%: MALE
Mean = 0.25001, Std Dev = 0.00108
Median = 0.25001
Mode = 0.25004
95% Lower CI = 0.24789
95% Upper CI = 0.25214
Count
Estimate

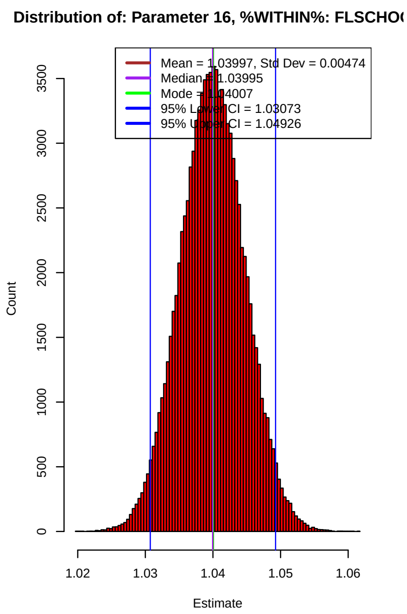

Distribution of: Parameter 11, %WITHIN%: IMMI1GE
Mean = 0.02821, Std Dev = 0.00012
Median = 0.02821
Mode = 0.02822
95% Lower CI = 0.02797
95% Upper CI = 0.02846
Count
Estimate

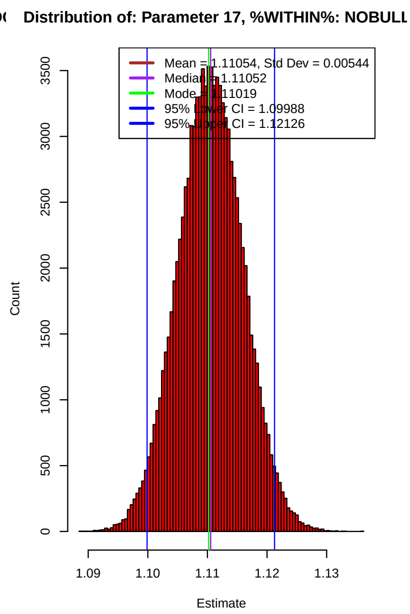

Distribution of: Parameter 12, %WITHIN%: IMMI2GE
Mean = 0.04061, Std Dev = 0.00018
Median = 0.04061
Mode = 0.04060
95% Lower CI = 0.04026
95% Upper CI = 0.04096
Count
Estimate

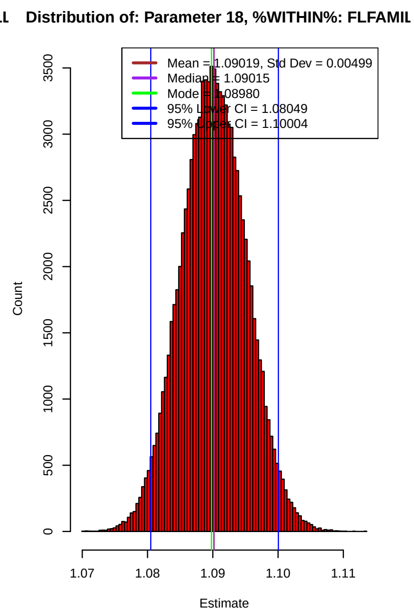

Distribution of: Parameter 13, %WITHIN%: ESCS
Mean = 1.18275, Std Dev = 0.00516
Median = 1.18274
Mode = 1.18282
95% Lower CI = 1.17265
95% Upper CI = 1.19291
Count
Estimate

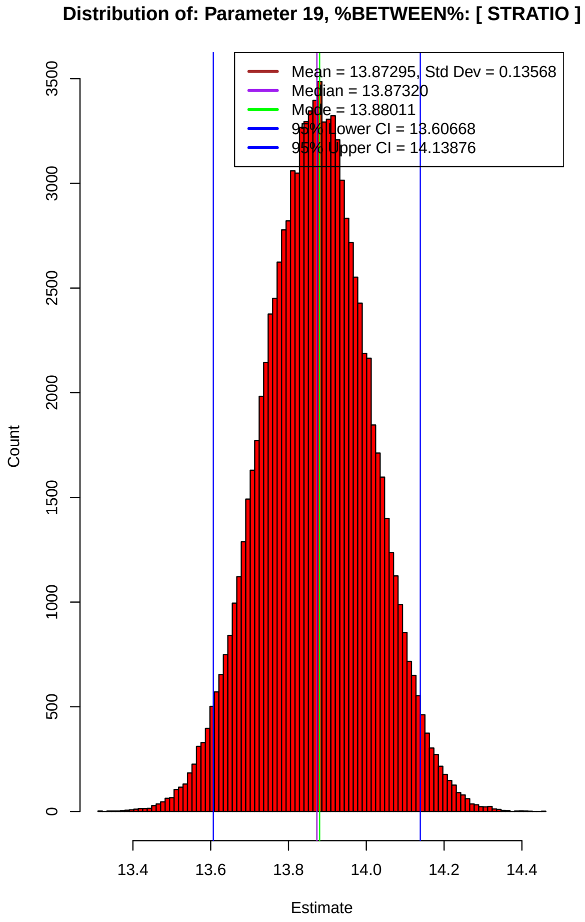

Distribution of: Parameter 14, %WITHIN%: FCFMLRT
Mean = 29.75375, Std Dev = 0.13273
Median = 29.75334
Mode = 29.77229
95% Lower CI = 29.49523
95% Upper CI = 30.01564
Count
Estimate

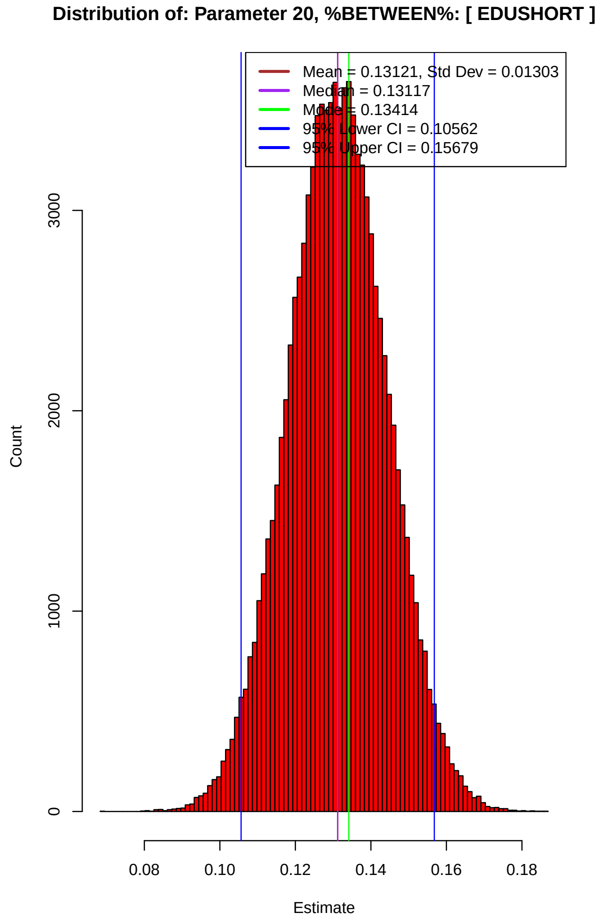

Distribution of: Parameter 15, %WITHIN%: FLCONF
Mean = 1.03417, Std Dev = 0.00488
Median = 1.03413
Mode = 1.03400
95% Lower CI = 1.02464
95% Upper CI = 1.04381
Count
Estimate

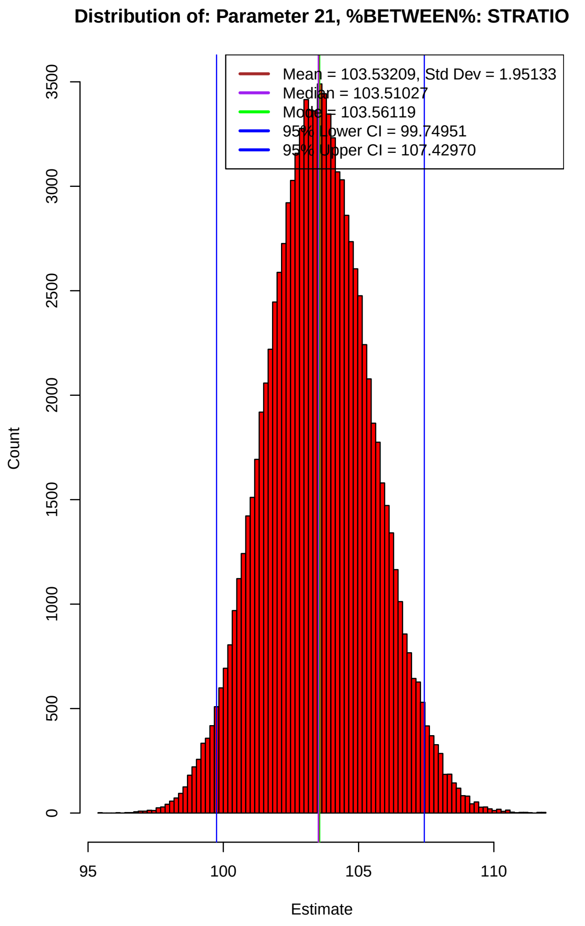

Distribution of: Parameter 16, %WITHIN%: FLSCHO
Mean = 1.03997, Std Dev = 0.00474
Median = 1.03995
Mode = 1.04007
95% Lower CI = 1.03073
95% Upper CI = 1.04926
Count
Estimate

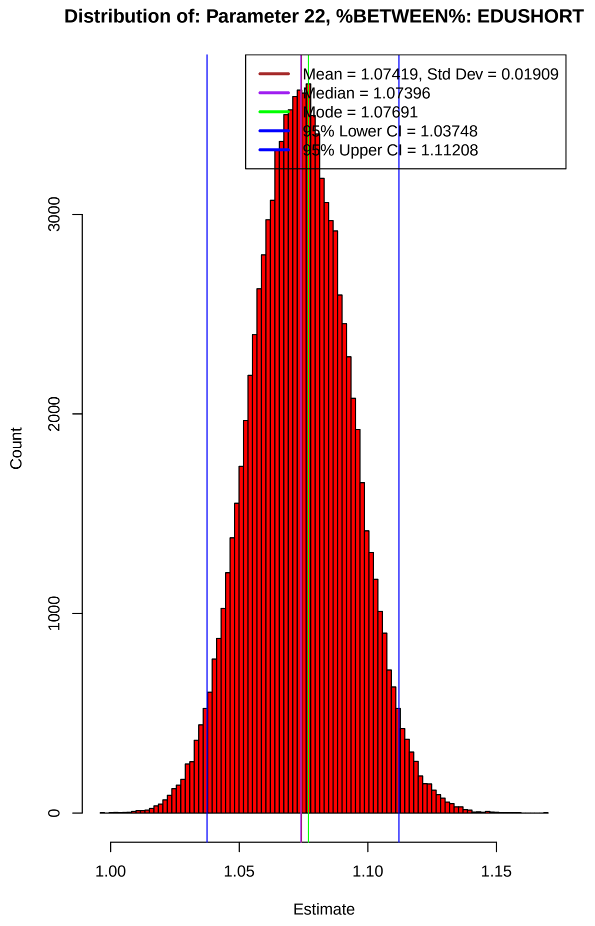

Distribution of: Parameter 17, %WITHIN%: NOBULL
Mean = 1.11054, Std Dev = 0.00544
Median = 1.11052
Mode = 1.11019
95% Lower CI = 1.09988
95% Upper CI = 1.12126
Count
Estimate

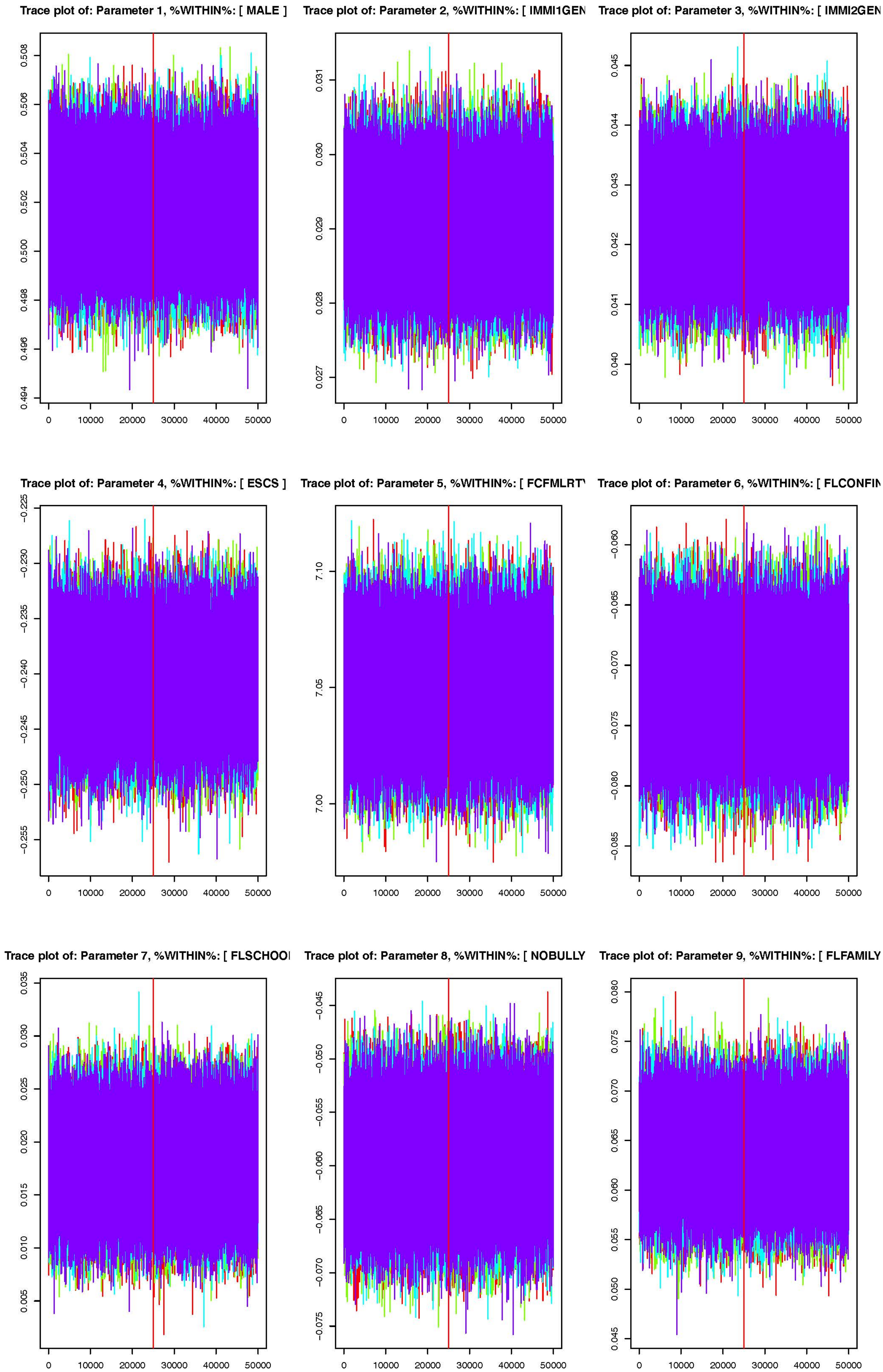

Distribution of: Parameter 18, %WITHIN%: FLFAMIL
Mean = 1.09019, Std Dev = 0.00499
Median = 1.09015
Mode = 1.08980
95% Lower CI = 1.08049
95% Upper CI = 1.10004
Count
Estimate

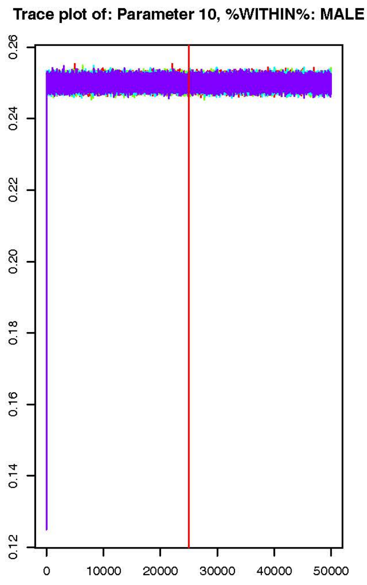

Distribution of: Parameter 19, %BETWEEN%: [STRATIO]
Mean = 13.87295, Std Dev = 0.13568
Median = 13.87320
Mode = 13.88011
95% Lower CI = 13.60668
95% Upper CI = 14.13876
Count
Estimate

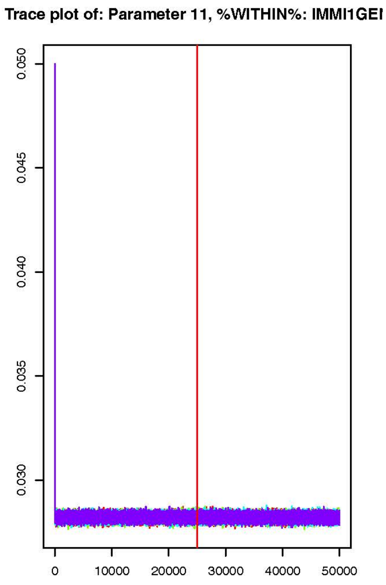

Distribution of: Parameter 20, %BETWEEN%: [EDUSHORT]
Mean = 0.13121, Std Dev = 0.01303
Median = 0.13117
Mode = 0.13414
95% Lower CI = 0.10562
95% Upper CI = 0.15679
Count
Estimate

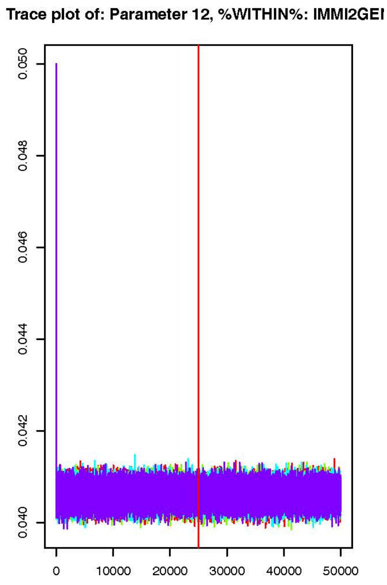

Distribution of: Parameter 21, %BETWEEN%: STRATIO
Mean = 103.53209, Std Dev = 1.95133
Median = 103.51027
Mode = 103.56119
95% Lower CI = 99.74951
95% Upper CI = 107.42970
Count
Estimate

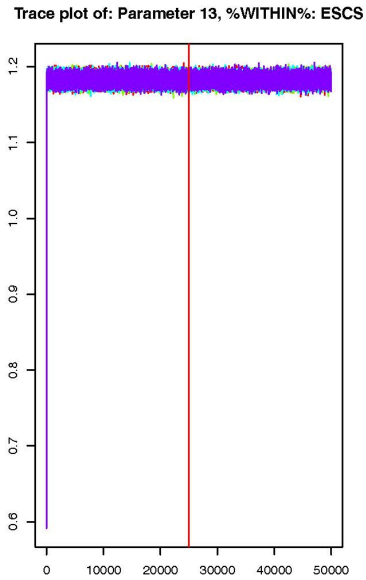

Distribution of: Parameter 22, %BETWEEN%: EDUSHORT
Mean = 1.07419, Std Dev = 0.01909
Median = 1.07396
Mode = 1.07691
95% Lower CI = 1.03748
95% Upper CI = 1.11208
Count
Estimate

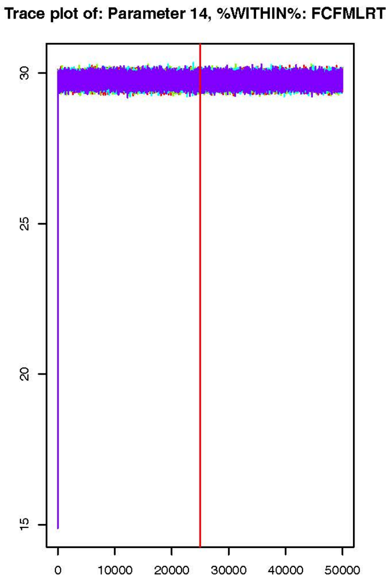

Trace plot of: Parameter 1, %WITHIN%: [MALE]
Trace plot of: Parameter 2, %WITHIN%: [IMMI1GEN
Trace plot of: Parameter 3, %WITHIN%: [IMMI2GEN
Trace plot of: Parameter 4, %WITHIN%: [ESCS]
Trace plot of: Parameter 5, %WITHIN%: [FCFMLRTY
Trace plot of: Parameter 6, %WITHIN%: [FLCONFIN
Trace plot of: Parameter 7, %WITHIN%: [FLSCHOOL
Trace plot of: Parameter 8, %WITHIN%: [NOBULLY
Trace plot of: Parameter 9, %WITHIN%: [FLFAMILY
0
10000
20000
30000
40000
50000

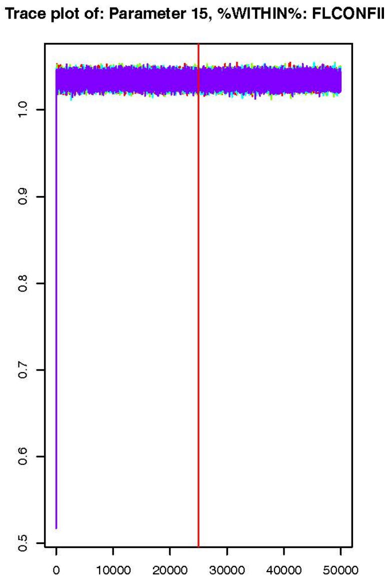

Trace plot of: Parameter 10, %WITHIN%: MALE

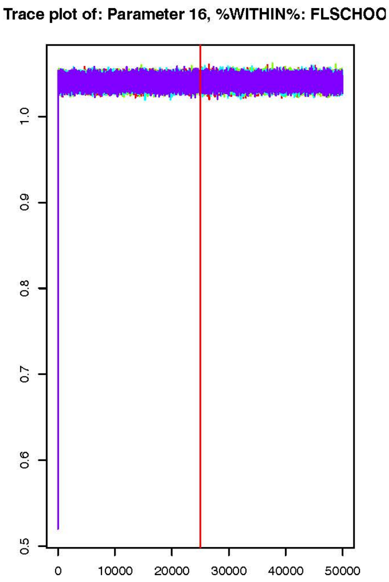

Trace plot of: Parameter 11, %WITHIN%: IMMI1GEI

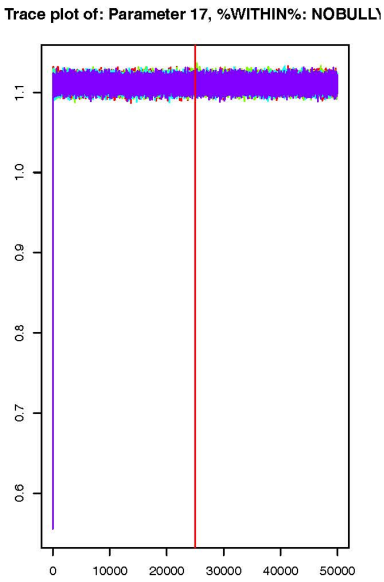

Trace plot of: Parameter 12, %WITHIN%: IMMI2GEI

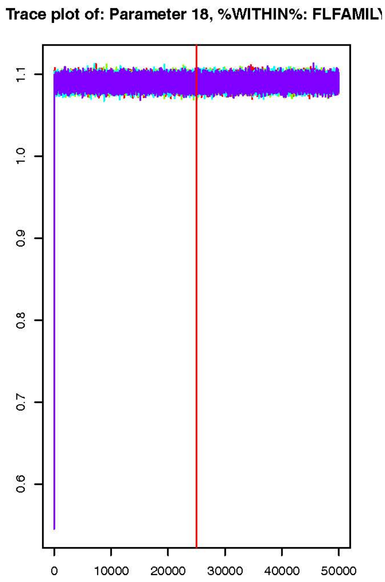

Trace plot of: Parameter 13, %WITHIN%: ESCS


Trace plot of: Parameter 14, %WITHIN%: FCFMLRT

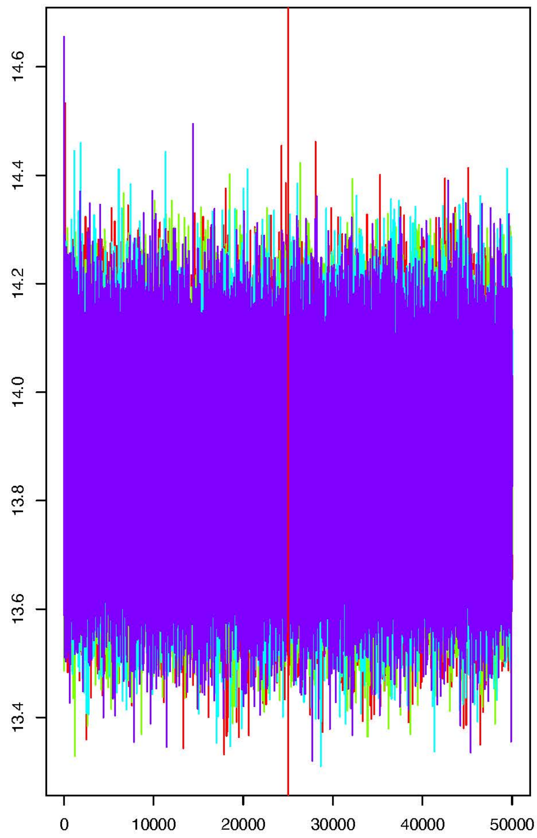

Trace plot of: Parameter 15, %WITHIN%: FLCONFII


Trace plot of: Parameter 16, %WITHIN%: FLSCHOO

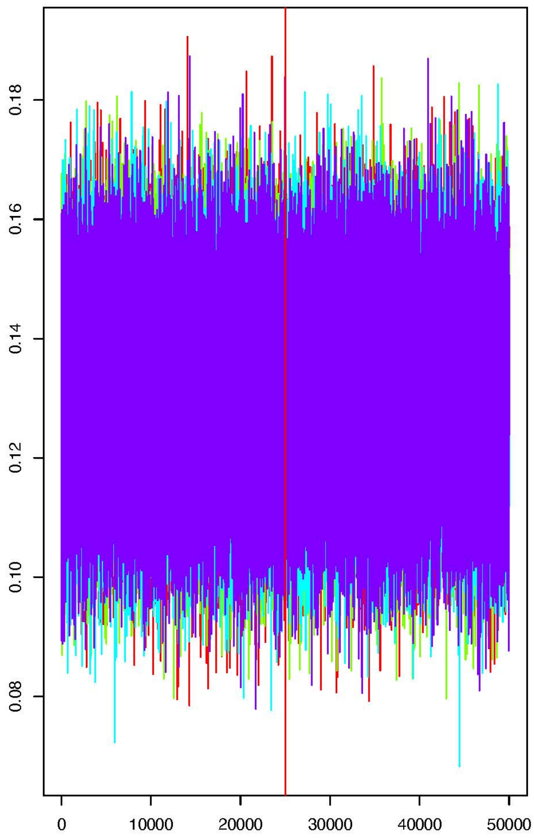

Trace plot of: Parameter 17, %WITHIN%: NOBULLY


Trace plot of: Parameter 18, %WITHIN%: FLFAMILY

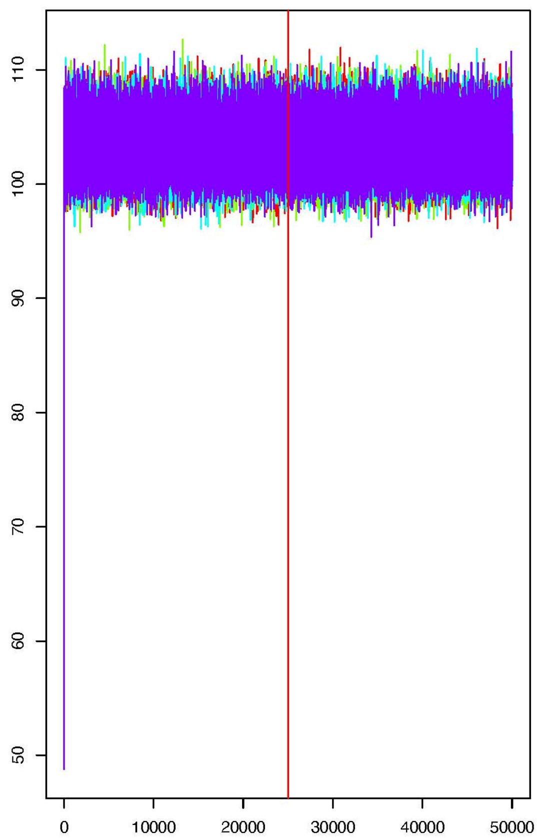
Trace plot of: Parameter 19, %BETWEEN%: [STRATIO]



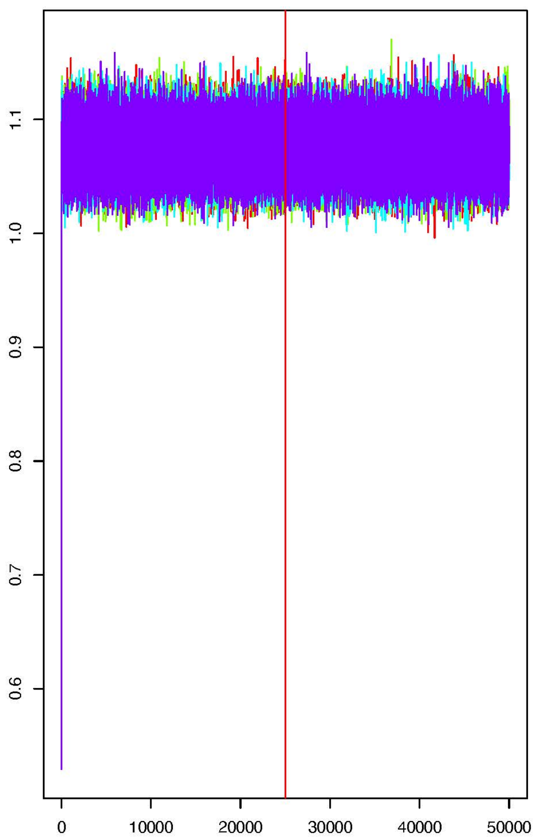
Trace plot of: Parameter 20, %BETWEEN%: [EDUSHORT]

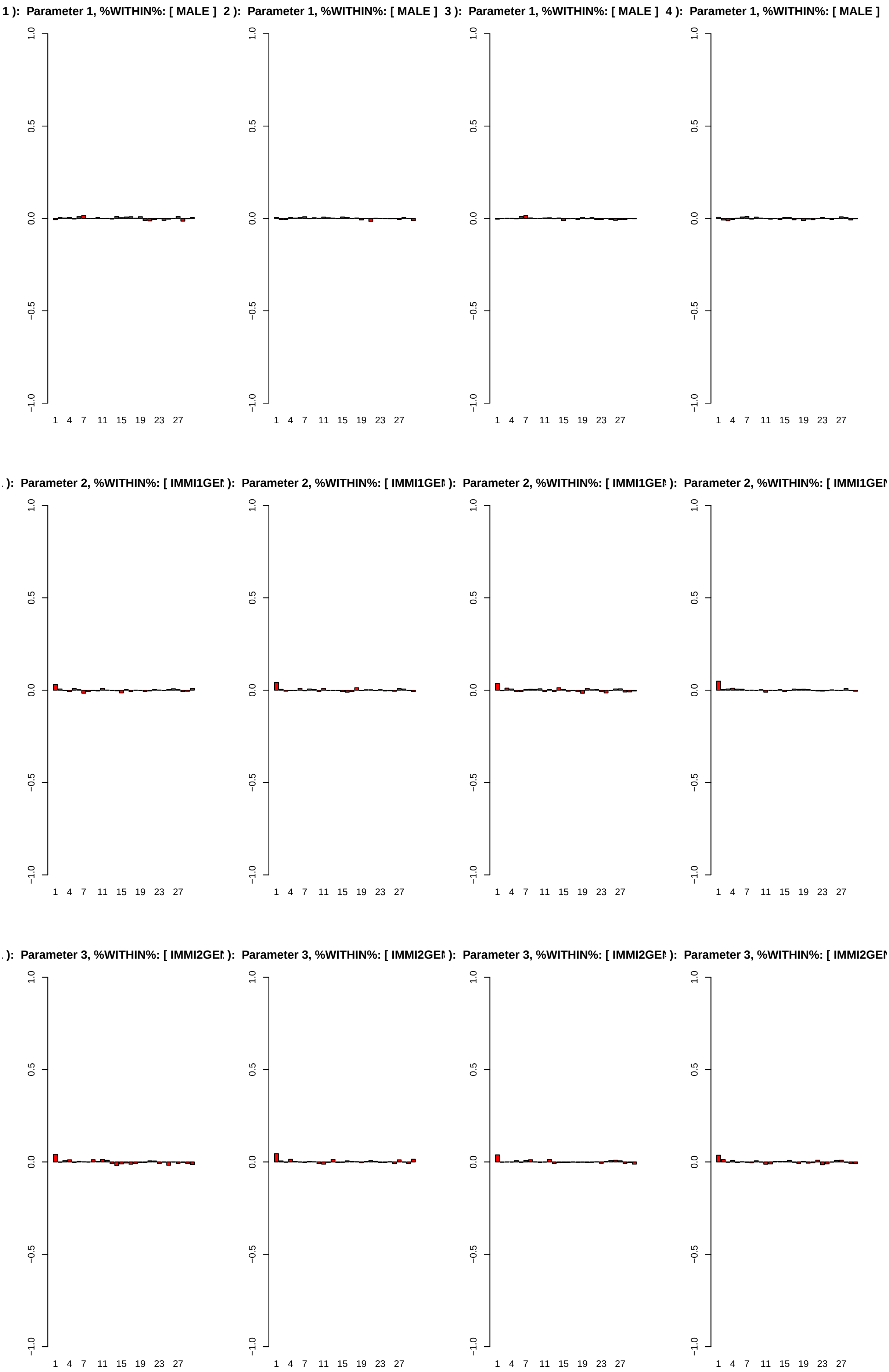


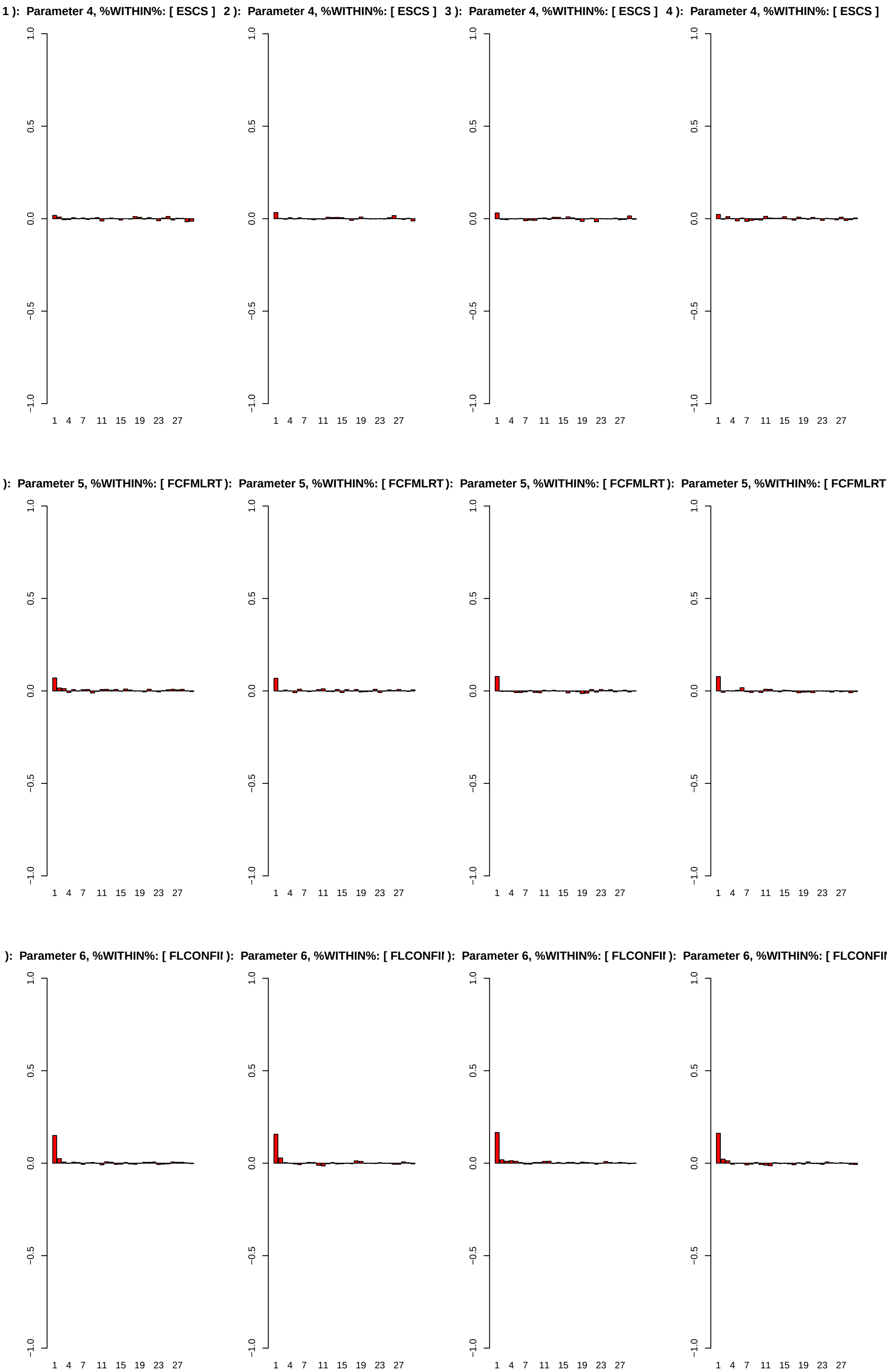
Trace plot of: Parameter 21, %BETWEEN%: STRATIO

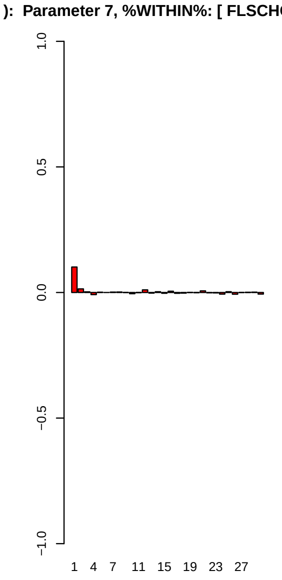
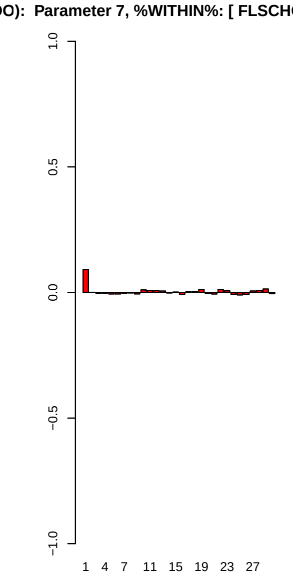
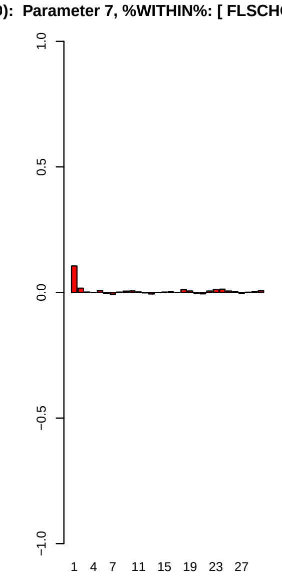
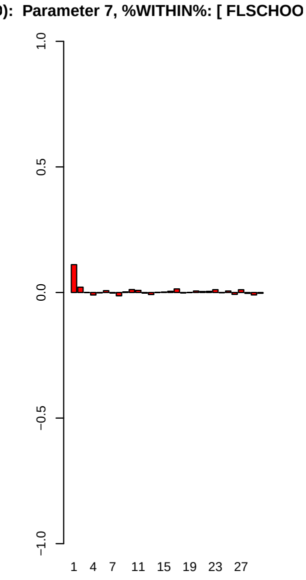
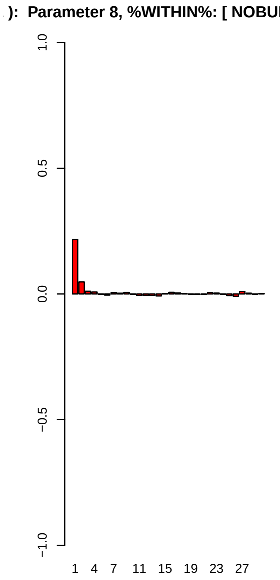
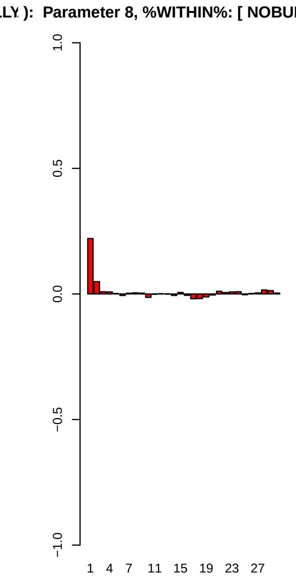
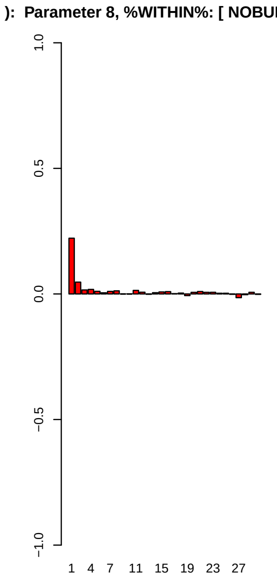
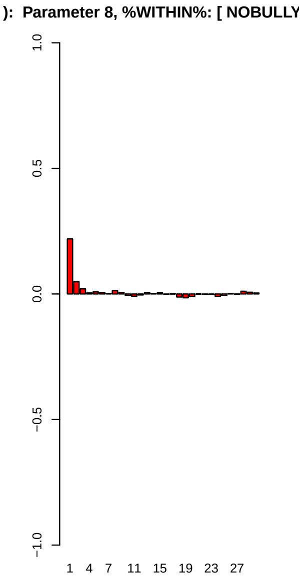
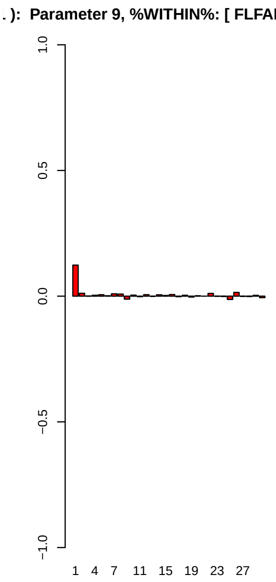
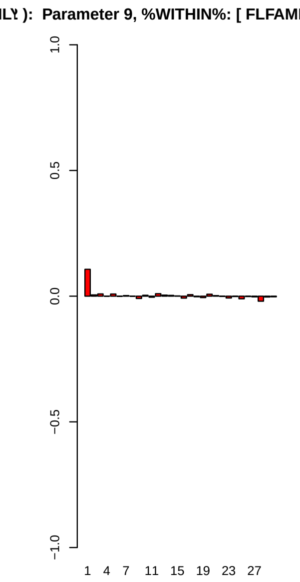
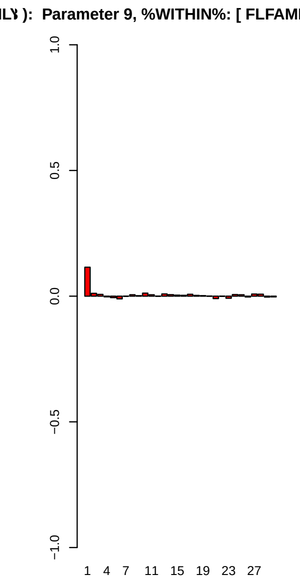
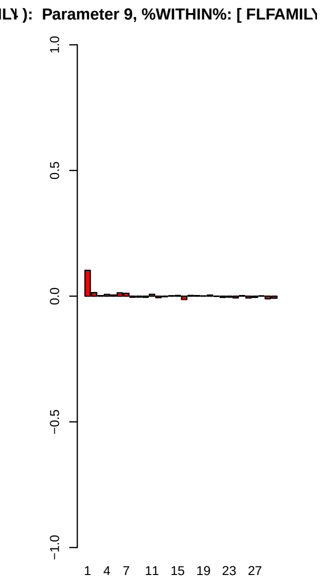


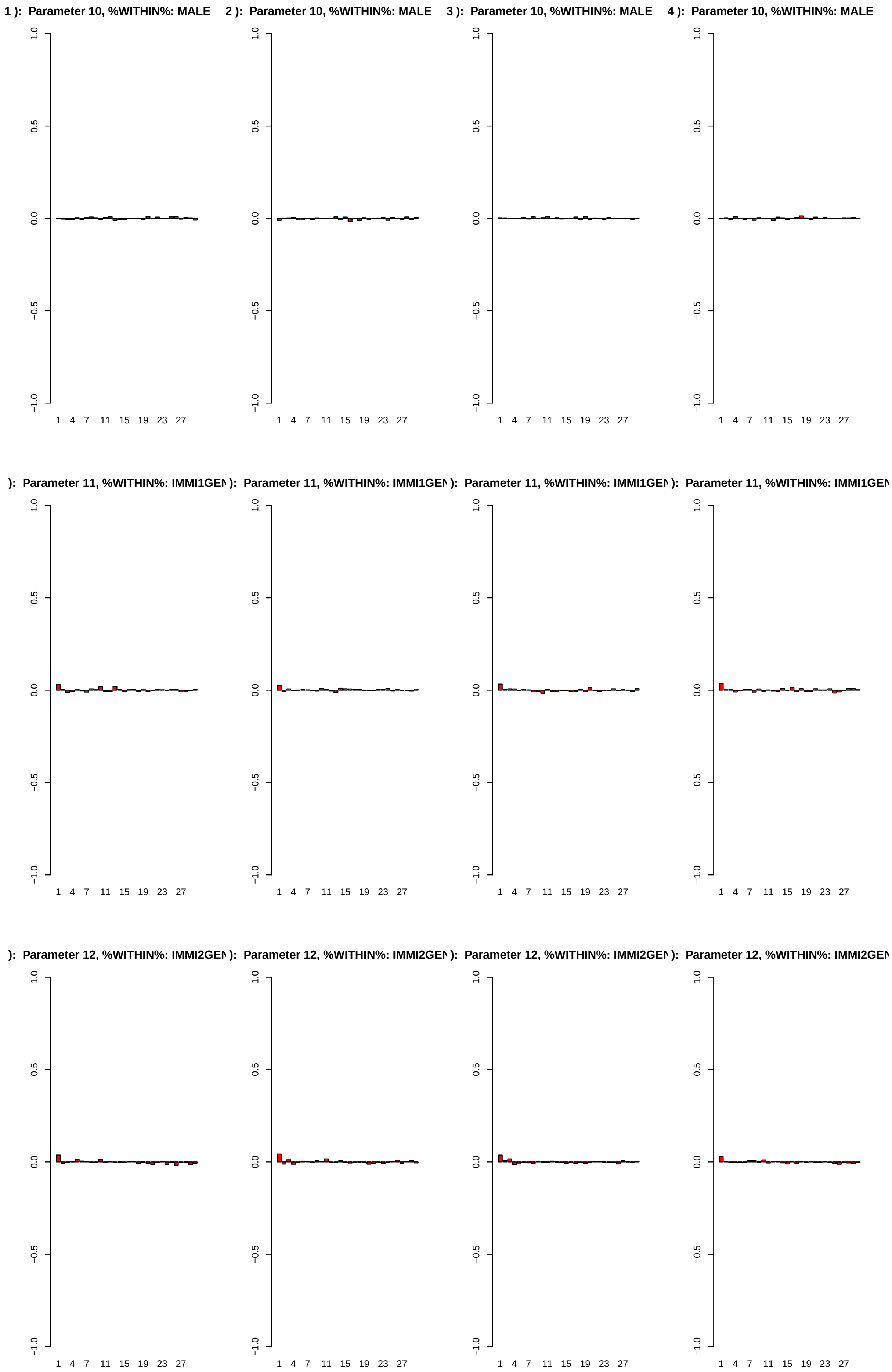
Trace plot of: Parameter 22, %BETWEEN%: EDUSHORT



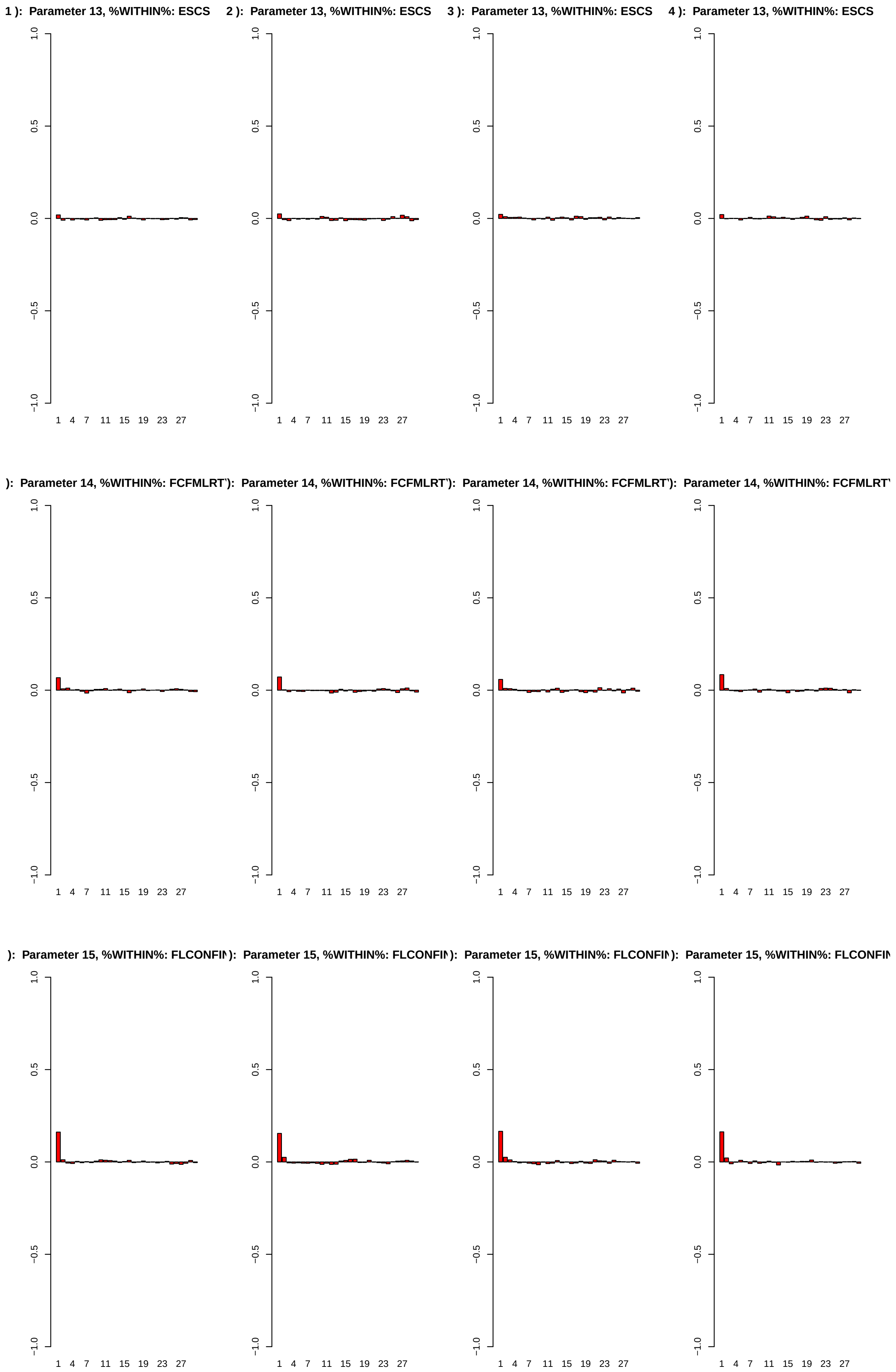

1): Parameter 1, %WITHIN%: [MALE]
2): Parameter 1, %WITHIN%: [MALE]
3): Parameter 1, %WITHIN%: [MALE]
4): Parameter 1, %WITHIN%: [MALE]
): Parameter 2, %WITHIN%: [IMMI1GE
): Parameter 2, %WITHIN%: [IMMI1GE
): Parameter 2, %WITHIN%: [IMMI1GE
): Parameter 2, %WITHIN%: [IMMI1GE
): Parameter 3, %WITHIN%: [IMMI2GE
): Parameter 3, %WITHIN%: [IMMI2GE
): Parameter 3, %WITHIN%: [IMMI2GE
): Parameter 3, %WITHIN%: [IMMI2GE
1.0
0.5
0.0
−0.5
−1.0
1 4 7 11 15 19 23 27


1): Parameter 4, %WITHIN%: [ESCS]
2): Parameter 4, %WITHIN%: [ESCS]
3): Parameter 4, %WITHIN%: [ESCS]
4): Parameter 4, %WITHIN%: [ESCS]
): Parameter 5, %WITHIN%: [FCFMLRT
): Parameter 5, %WITHIN%: [FCFMLRT
): Parameter 5, %WITHIN%: [FCFMLRT
): Parameter 5, %WITHIN%: [FCFMLRT
): Parameter 6, %WITHIN%: [FLCONFI
): Parameter 6, %WITHIN%: [FLCONFI
): Parameter 6, %WITHIN%: [FLCONFI
): Parameter 6, %WITHIN%: [FLCONFI
1.0
0.5
0.0
−0.5
−1.0
1 4 7 11 15 19 23 27

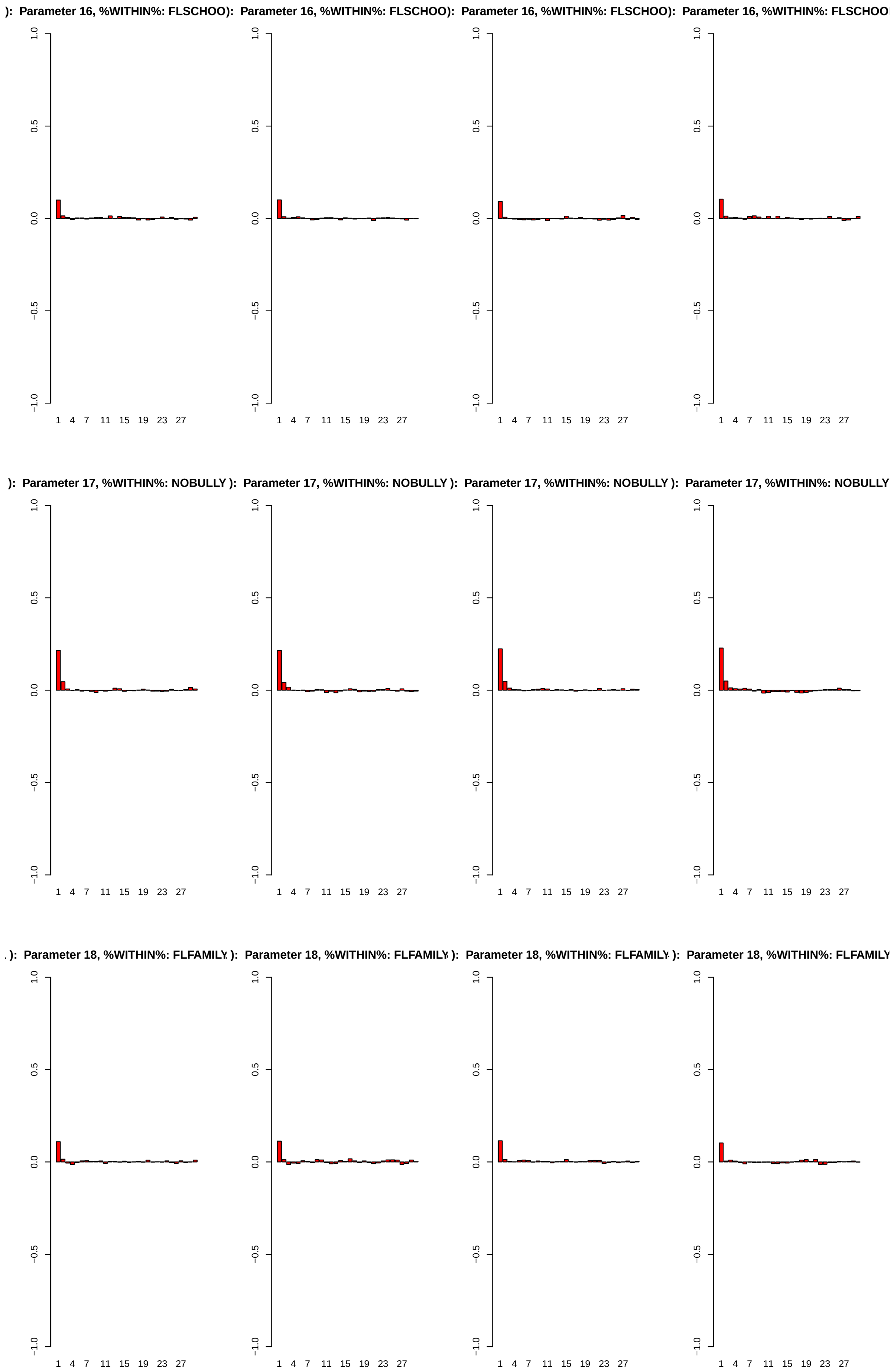

): Parameter 7, %WITHIN%: [FLSCHOO
1.0
0.5
0.0
−0.5
−1.0
1 4 7 11 15 19 23 27

): Parameter 7, %WITHIN%: [FLSCHOO
1.0
0.5
0.0
−0.5
−1.0
1 4 7 11 15 19 23 27

): Parameter 7, %WITHIN%: [FLSCHOO
1.0
0.5
0.0
−0.5
−1.0
1 4 7 11 15 19 23 27

): Parameter 7, %WITHIN%: [FLSCHOO
1.0
0.5
0.0
−0.5
−1.0
1 4 7 11 15 19 23 27

.): Parameter 8, %WITHIN%: [NOBULLY
1.0
0.5
0.0
−0.5
−1.0
1 4 7 11 15 19 23 27

): Parameter 8, %WITHIN%: [NOBULLY
1.0
0.5
0.0
−0.5
−1.0
1 4 7 11 15 19 23 27

): Parameter 8, %WITHIN%: [NOBULLY
1.0
0.5
0.0
−0.5
−1.0
1 4 7 11 15 19 23 27

): Parameter 8, %WITHIN%: [NOBULLY
1.0
0.5
0.0
−0.5
−1.0
1 4 7 11 15 19 23 27

.): Parameter 9, %WITHIN%: [FLFAMILY
1.0
0.5
0.0
−0.5
−1.0
1 4 7 11 15 19 23 27

): Parameter 9, %WITHIN%: [FLFAMILY
1.0
0.5
0.0
−0.5
−1.0
1 4 7 11 15 19 23 27

): Parameter 9, %WITHIN%: [FLFAMILY
1.0
0.5
0.0
−0.5
−1.0
1 4 7 11 15 19 23 27

): Parameter 9, %WITHIN%: [FLFAMILY
1.0
0.5
0.0
−0.5
−1.0
1 4 7 11 15 19 23 27

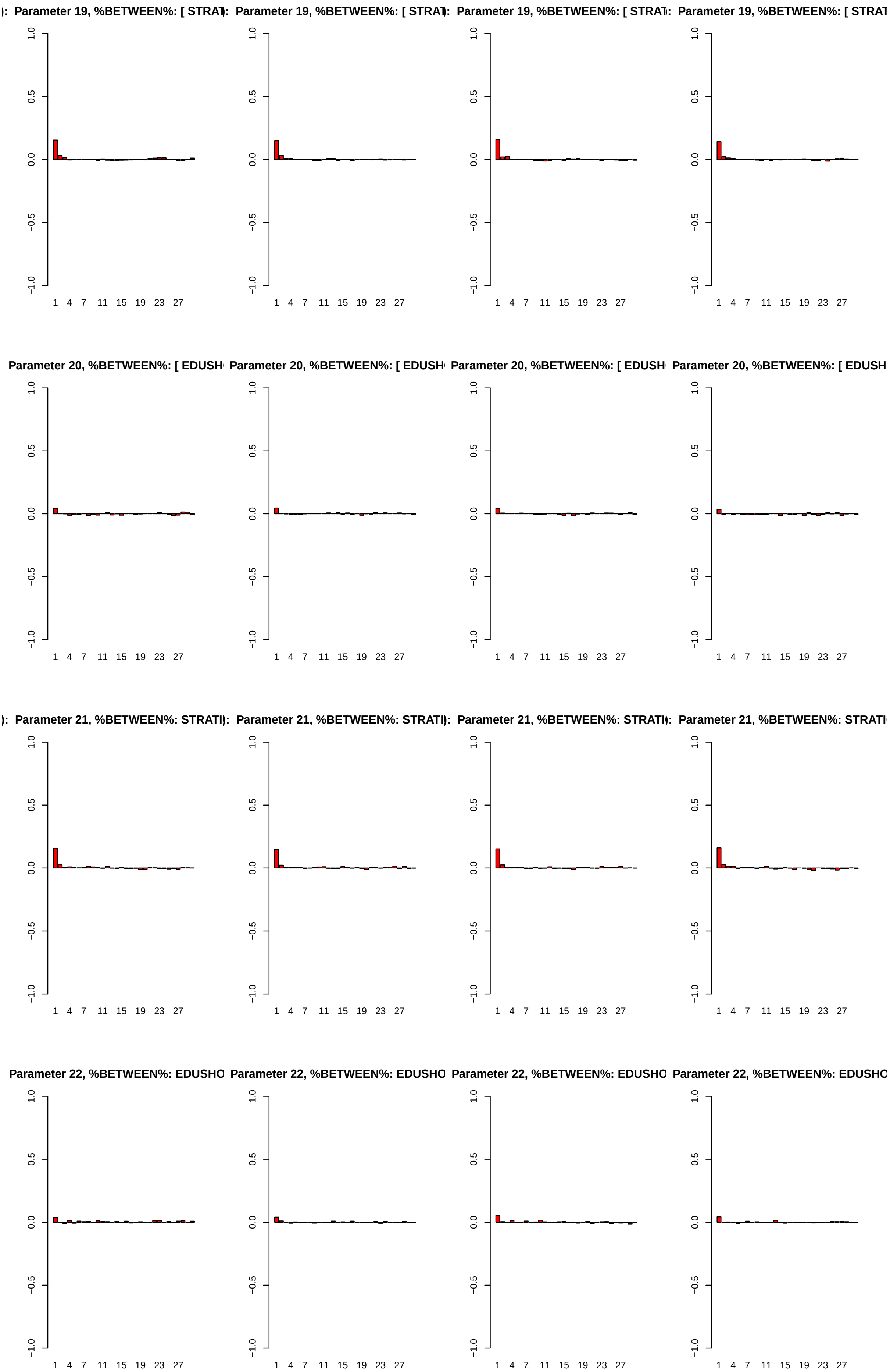


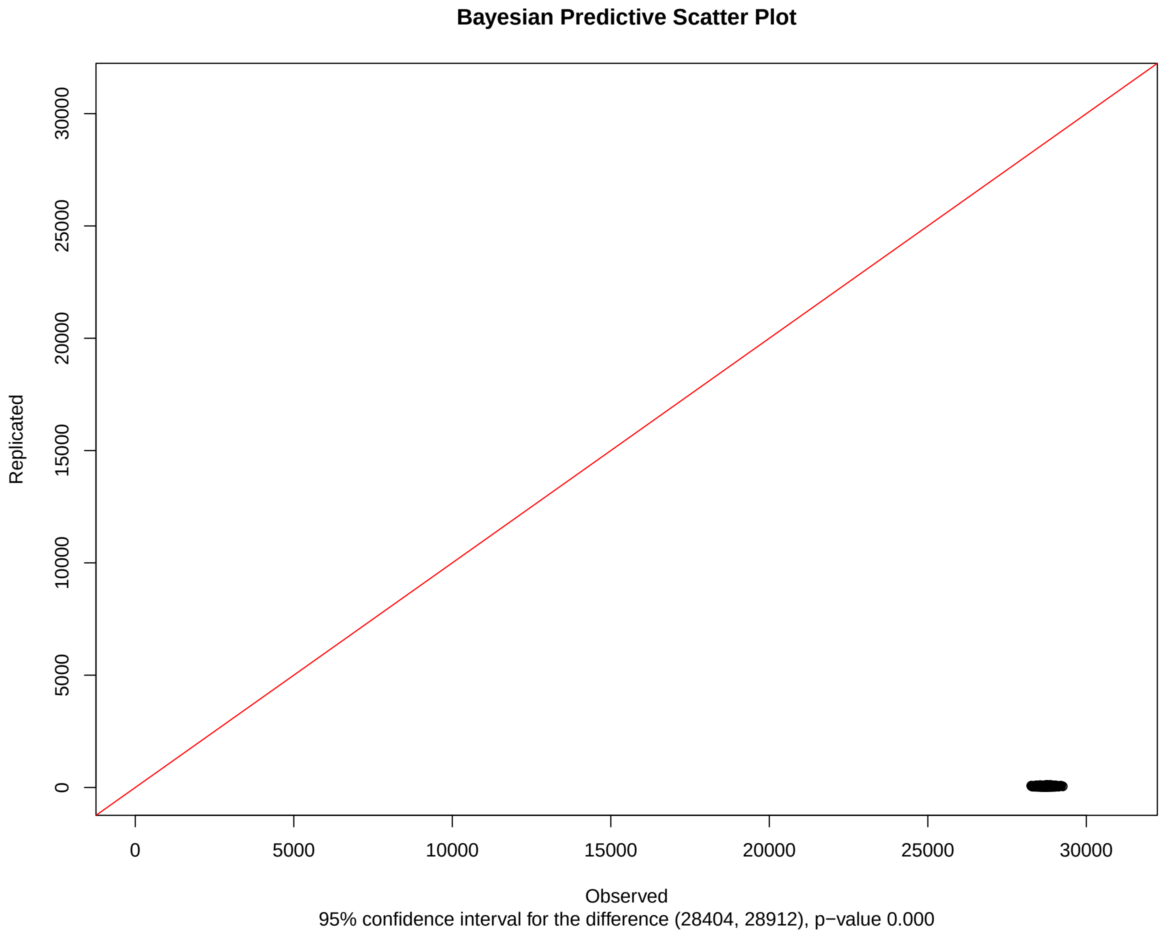
1): Parameter 10, %WITHIN%: MALE
2): Parameter 10, %WITHIN%: MALE
3): Parameter 10, %WITHIN%: MALE
4): Parameter 10, %WITHIN%: MALE
): Parameter 11, %WITHIN%: IMMI1GEN
): Parameter 11, %WITHIN%: IMMI1GEN
): Parameter 11, %WITHIN%: IMMI1GEN
): Parameter 11, %WITHIN%: IMMI1GEN
): Parameter 12, %WITHIN%: IMMI2GEN
): Parameter 12, %WITHIN%: IMMI2GEN
): Parameter 12, %WITHIN%: IMMI2GEN
): Parameter 12, %WITHIN%: IMMI2GEN
1.0
0.5
0.0
−0.5
−1.0
1 4 7 11 15 19 23 27

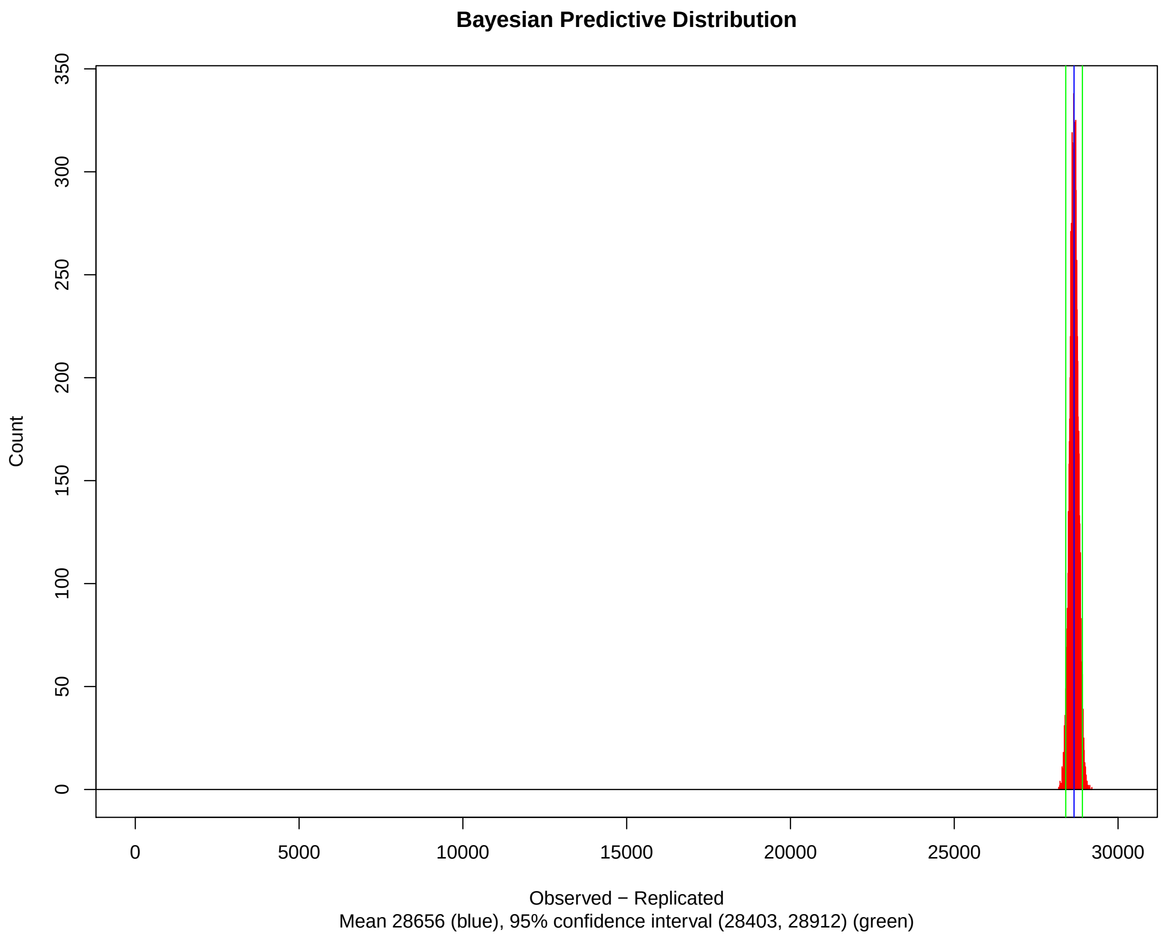


1): Parameter 13, %WITHIN%: ESCS
2): Parameter 13, %WITHIN%: ESCS
3): Parameter 13, %WITHIN%: ESCS
4): Parameter 13, %WITHIN%: ESCS
): Parameter 14, %WITHIN%: FCFMLRT'
): Parameter 14, %WITHIN%: FCFMLRT'
): Parameter 14, %WITHIN%: FCFMLRT'
): Parameter 14, %WITHIN%: FCFMLRT'
): Parameter 15, %WITHIN%: FLCONFII
): Parameter 15, %WITHIN%: FLCONFII
): Parameter 15, %WITHIN%: FLCONFII
): Parameter 15, %WITHIN%: FLCONFIN
1.0
0.5
0.0
−0.5
−1.0
1 4 7 11 15 19 23 27


): Parameter 16, %WITHIN%: FLSCHOO): Parameter 16, %WITHIN%: FLSCHOO): Parameter 16, %WITHIN%: FLSCHOO): Parameter 16, %WITHIN%: FLSCHOO
): Parameter 17, %WITHIN%: NOBULLY): Parameter 17, %WITHIN%: NOBULLY): Parameter 17, %WITHIN%: NOBULLY): Parameter 17, %WITHIN%: NOBULLY
): Parameter 18, %WITHIN%: FLFAMILY): Parameter 18, %WITHIN%: FLFAMILY): Parameter 18, %WITHIN%: FLFAMILY): Parameter 18, %WITHIN%: FLFAMILY
1.0
0.5
0.0
−0.5
−1.0
1 4 7 11 15 19 23 27


Parameter 19, %BETWEEN%: [STRAT
Parameter 19, %BETWEEN%: [STRAT
Parameter 19, %BETWEEN%: [STRAT
Parameter 19, %BETWEEN%: [STRAT
Parameter 20, %BETWEEN%: [EDUSH
Parameter 20, %BETWEEN%: [EDUSH
Parameter 20, %BETWEEN%: [EDUSH
Parameter 20, %BETWEEN%: [EDUSH
Parameter 21, %BETWEEN%: STRATI
Parameter 21, %BETWEEN%: STRATI
Parameter 21, %BETWEEN%: STRATI
Parameter 21, %BETWEEN%: STRATI
Parameter 22, %BETWEEN%: EDUSHO
Parameter 22, %BETWEEN%: EDUSHO
Parameter 22, %BETWEEN%: EDUSHO
Parameter 22, %BETWEEN%: EDUSHO
1.0
0.5
0.0
−0.5
−1.0
1 4 7 11 15 19 23 27


Bayesian Predictive Scatter Plot
Replicated
Observed
95% confidence interval for the difference (28404, 28912), p−value 0.000


Bayesian Predictive Distribution
Count
Observed − Replicated
Mean 28656 (blue), 95% confidence interval (28403, 28912) (green)

B.3 MSEM Analysis Code

B.3.1 *Mplus* Input

```
TITLE:
    Two-level structured model for all countries

DATA:
    file = "~/implist.dat";

    type = imputation;


VARIABLE:
    names =
        FKI CNTRYID CNTSCHID W_STU
        MATH READ FLIT                       ! PISA achievement variables
        MALE IMMI1GEN IMMI2GEN ESCS          ! Demographic info
        FCFMLRTY FLCONFIN                    ! Affect
        FLSCHOOL                             !   Lat var "Academic"
        NOBULLY                              !   Lat var "Safety"
        FLFAMILY                             !   Lat var "Community"
        W_SCH STRATIO                        ! School character
        EDUSHORT                             !   Lat var "inst. env."
        ;

    usevar =
        FLIT                                 ! PISA achievement variables
        FLSCHOOL                             !   Lat var "Academic"
        NOBULLY                              !   Lat var "Safety"
        FLFAMILY                             !   Lat var "Community"
        EDUSHORT STRATIO                     !   Lat var "Inst env"
        FCFMLRTY FLCONFIN                    ! Affect
        MALE IMMI1GEN IMMI2GEN ESCS          ! Demographic info
        ;

    ! Vars that ONLY appear in L1
    within =
        FCFMLRTY FLCONFIN                    ! Affective vars
        MALE IMMI1GEN IMMI2GEN ESCS          ! L1 control vars
        ;

    ! Vars that ONLY appear in L2
    between =
        EDUSHORT STRATIO                     ! L2: school
        ;

    weight = W_STU;                          ! Student weight
    wtscale = cluster;                       ! Scale L1 weight to cluster size
    bweight = W_SCH;                         ! School weight
    bwtscale = sample;                       ! Scale L2 weight to sample

    cluster = CNTSCHID;                      ! Cluster by school ID


ANALYSIS:
    processors = 64;

    type = twolevel;


MODEL:

%Within%                                     ! === L1: Student-level ===

    ! Save the variances of L1 FLSCHOOL, FLFAMILY, NOBULLY and FLIT
    FLSCHOOL (va_w);                         ! variance of academic (within)
    FLFAMILY (vc_w);                         ! variance of community (within)
    NOBULLY (vs_w);                          ! variance of safety (within)
    FLIT (vf_w);                             ! variance of FLIT (within)
```

```
    ! Indirect pathways (1st half): school climate vars onto mediators
        ! Onto mediator FAMILIARITY
    FCFMLRTY on
        FLSCHOOL FLFAMILY NOBULLY
        ESCS IMMI2GEN MALE
        ;
        ! Onto mediator CONFIDENCE
    FLCONFIN on
        FLSCHOOL FLFAMILY NOBULLY
        ESCS IMMI2GEN MALE
        ;

    ! Total effect
    FLIT on
        ! Indirect pathways (2nd half): affective vars onto financial literacy
        FCFMLRTY FLCONFIN
        ! Direct pathways: school climate vars onto financial literacy
        FLSCHOOL (a_w)                      ! academic_within
        FLFAMILY (c_w)                      ! community_within
        NOBULLY (s_w)                       ! safety_within
        ! Demographic vars
        ESCS IMMI1GEN
        ;

    ! Covariances

        ! Between school climate vars
    FLSCHOOL with FLFAMILY;
    FLFAMILY with NOBULLY;
    FLSCHOOL with NOBULLY;

        ! Between mediators
    FCFMLRTY with FLCONFIN;

        ! SES with school climate vars
    ESCS with FLSCHOOL FLFAMILY;
        ! SES with demographic vars
    ESCS with IMMI1GEN IMMI2GEN;

%Between%                                   ! === L2: School-level ===

    ! Save the variances of L1 FLSCHOOL, FLFAMILY and NOBULLY
    FLSCHOOL (va_b);                        ! variance of academic (between)
    FLFAMILY (vc_b);                        ! variance of community (between)
    NOBULLY (vs_b);                         ! variance of safety (between)
    FLIT (vf_b);                            ! variance of FLIT (between)

    FLIT on                                 ! School climate variables
        FLSCHOOL (a_b)                          ! academic_between
        FLFAMILY (c_b)                          ! community_between
        NOBULLY (s_b)                           ! safety_between
        EDUSHORT
        STRATIO                             ! Control: Student-teacher ratio
        ;

    ! Covariances

        ! Between school climate vars
    FLSCHOOL with FLFAMILY;
    FLFAMILY with NOBULLY;
    NOBULLY with EDUSHORT;

    FLSCHOOL with NOBULLY;
    FLFAMILY with EDUSHORT;

    FLSCHOOL with EDUSHORT;

    STRATIO with FLSCHOOL EDUSHORT;


MODEL INDIRECT:
```

```
    ! Indirect effects
    FLIT ind FLSCHOOL;
    FLIT ind FLFAMILY;
    FLIT ind NOBULLY;

    FLIT ind ESCS;
    FLIT ind IMMI2GEN;
    FLIT ind MALE;


MODEL CONSTRAINT:

    ! Save non-standardised contextual effects
    new(ctx_a);
    ctx_a = a_b - a_w;
    new(ctx_c);
    ctx_c = c_b - c_w;
    new(ctx_s);
    ctx_s = s_b - s_w;

    ! Standardise contextual effects
    new(ctx_a_st);
    ctx_a_st = ctx_a*(sqrt(va_b)/sqrt(va_b*a_b**2+vf_b+va_w*a_w**2+vf_w));
    new(ctx_c_st);
    ctx_c_st = ctx_c*(sqrt(vc_b)/sqrt(vc_b*c_b**2+vf_b+vc_w*c_w**2+vf_w));
    new(ctx_s_st);
    ctx_s_st = ctx_s*(sqrt(vs_b)/sqrt(vs_b*s_b**2+vf_b+vs_w*s_w**2+vf_w));

    ! Compute effect sizes (EF)
    new(es1_a);
    es1_a = ctx_a*(2*sqrt(va_b)/sqrt(vf_w));
    new(es1_c);
    es1_c = ctx_c*(2*sqrt(vc_b)/sqrt(vf_w));
    new(es1_s);
    es1_s = ctx_s*(2*sqrt(vs_b)/sqrt(vf_w));

    new(es2_a);
    es2_a = ctx_a*(2*sqrt(va_b)/sqrt(va_w*a_w**2+vf_w));
    new(es2_c);
    es2_c = ctx_c*(2*sqrt(vc_b)/sqrt(vc_w*c_w**2+vf_w));
    new(es2_s);
    es2_s = ctx_s*(2*sqrt(vs_b)/sqrt(vs_w*s_w**2+vf_w));

    new(es3_a);
    es3_a = ctx_a*(2*sqrt(va_b)/sqrt(va_b*a_b**2+va_w*a_w**2+vf_w));
    new(es3_c);
    es3_c = ctx_c*(2*sqrt(vc_b)/sqrt(vc_b*c_b**2+vc_w*c_w**2+vf_w));
    new(es3_s);
    es3_s = ctx_s*(2*sqrt(vs_b)/sqrt(vs_b*s_b**2+vs_w*s_w**2+vf_w));


OUTPUT:
    stdyx                                       ! Fully standardised solution
    ;
```

B.3.2 Selected *Mplus* Output

```
R-SQUARE

Within Level

    Observed                                            Two-Tailed    Rate of
    Variable         Estimate       S.E.  Est./S.E.     P-Value      Missing

    FLIT                0.122      0.009     12.875       0.000        0.203
    FCFMLRTY            0.136      0.007     19.204       0.000        0.123
    FLCONFIN            0.077      0.005     14.908       0.000        0.141

Between Level

    Observed                                            Two-Tailed    Rate of
```

```
Variable          Estimate        S.E.  Est./S.E.    P-Value    Missing

FLIT                 0.477       0.038     12.469      0.000      0.062
```

Appendix C

Selected Work: Item Response Theory (MAE4120) Final Assignment

Sex Differences in Norwegian Students' Perceptions of Teacher Unfairness

Tony C. A. Tan

Centre for Educational Measurement, University of Oslo

MAE4120 Item Response Theory

Dr Björn H. Andersson

20 April 2020

Sex Differences in Norwegian Students' Perceptions of Teacher Unfairness

Tony C. A. Tan
Centre for Educational Measurement, University of Oslo

Maintaining student well-being is paramount for educators. Adverse experiences from peers and/or teachers undermine students' physical and mental welfare, causing impediment to academic progress. The Programme for International Student Assessment (PISA) introduced a series of questions in its 2015 questionnaire to capture 15-year-old students' experience with peer bullying and teacher unfairness in school. Numerous studies have subsequently examined student response data for these two welfare measures, with various degrees of convergence towards some general agreements such as boys experiencing more episodes of bullying by peers and unfair teacher treatment. Prior research also reported difficulty in upholding statistical power when studying peer bullying questions due to the large proportion of "never or almost never" responses. An improved situation applies to teacher unfairness questions, and are therefore used for the current project. Sex difference is a concern from a methodological point of view because it threatens one of the main assumptions behind all item response theory (IRT) modelling—unidimensionality—by effectively introducing gender group as an extra dimension. Using Norwegian data, this project detects three of the six teacher unfairness items to display differential item functioning (DIF) over student sex. Missing data treatment using multiple imputation is also discussed.

Keywords: PISA, student well-being, teacher unfairness, sex differences, item response theory (IRT), differential item functioning (DIF)

Introduction

Students' well-being has direct and significant impacts on their academic performance as well as social development. By adolescence, students' perceptions of a positive school experience is found to be largely influenced by their subjective feeling of being treated fairly by teachers (Peter & Dalbert, 2010). In order to shed further insight into 15-year-olds' classroom experience, Programme for International Student Assessment (PISA) introduced a set of items in its 2015 testing cycle to specifically survey candidates' perceptions of their teachers' unfairness (OECD, 2017b).

Perceived Teacher Unfairness

Studies on perceptions of teacher unfairness largely build their theoretical foundation on organisation justice. By adapting the concept of organisational justice to the classroom setting, Chory-Assad (2002) was able to empirically separate undergraduate students' classroom justice ratings into three sub-scales and labelled them as `distributive`, `procedural` and `interactional` justice, in reference to the outcomes, the processes and the implementation of classroom fairness, respectively (Chory-Assad & Paulsel, 2004). Subsequent collaboration between Chory-Assad and colleagues generated more detailed catalogue within each category (Horan et al., 2010), leading to labels such as `grade`, `punishment` and `insensitive/rude`. Using these classification systems, Chen and Cui (2019) grouped 2015 PISA questions on teacher unfairness `ST039` (see Appendix A, an extract of OECD (2014)) into clusters, as summarised in Table 1.

It is interesting to observe that third-party verification can be made, at least in theory, to some of the six items under `ST039`. Student work can be re-marked by an independent examiner, for example, to verify students' claim of being graded more harshly by their teachers (`ST039Q02NA`); while it is less practical to independently attest whether a student is truly in the impression of being under-valued (`ST039Q03NA`). A column has therefore been augmented in Table 1 to capture such difference.

Sex Differences

Prior research has also suggested strong sex differences in school injustice perception. A French study by Lentillon et al. (2006) reported conclusive sex differences in students'

Correspondence: Tony C. A. Tan https://orcid.org/0000-0001-6632-3791, tctan@uio.no. Centre for Educational Measurement, Faculty of Educational Sciences, University of Oslo, Oslo, Norway

Table 1
Overview of `ST039` *Items*

Item Number	Chory-Assad (2004) Classification	Horan et al. (2010) Classification	Third-party Verifiable?	DIF Present?
ST039Q01NA	Distributive	instructor affect	Yes	Yes
ST039Q02NA	Distributive	grade	Yes	Yes
ST039Q03NA	Interactional	implied/stated stupidity	No	No
ST039Q04NA	Distributive	punishment	No	No
ST039Q05NA	Interactional	Insensitive/rude	Yes	Yes
ST039Q06NA	Interactional	Insensitive/rude	No	No

perceptions of grades but not in the aspect of teacher support; interestingly, it were boys, who on average received higher grades in physical education, that voiced to be more deprived in grades. A similar result was reported by Resh and Dalbert (2007) that both German and Israeli boys reported to have been disadvantaged in grades. OECD (2017a) also corroborates "boys were more likely than girls to report that their teachers do not treat them fairly." (p. 126), based on PISA 2015 results.

Differential Item Functioning (DIF)

Repeated reporting of sex difference in students' perceptions of teacher unfairness is concerning from a methodological point of view. One key assumption for the use of an IRT model to test item appropriateness is *unidimensionality*, which means that only one type of ability or achievement is being measured by the instrument (Walstad & Rebeck, 2017). Under such condition, responders possessing similar skills should yield similar probabilities of scoring correctly *irrespective of* sex (Holland & Wainer, 1993); if not, sex group is playing the role of another dimension in addition to the measurement tool (Roussos & Stout, 1996). Presence of sex differences threatens the validity of IRT modelling, with the potential of misleading researchers into making wrong decisions about the individuals.

DIF shall be employed to investigate the presence of biases resultant from multiple group effect (Angoff, 1993). By comparing the relative fit of two IRT models using a likelihood ratio test (Thissen et al., 1993), a DIF procedure is able to detect any statistical difference between a restricted model, where all groups are forced to share the same parameters, and an unrestricted model, where they are free to put on their own values. Large χ^2 statistics suggest potential DIF patterns (Bock & Zimowski, 1997).

Research Questions

Although larger sample sizes generally strengthen statistical power validity (Shadish et al., 2001), cross-country cultural differences may distort response patterns. Resultantly, a single country sample design is preferred over a multi-country dataset. As a research institute located in Oslo, this project wishes to investigate:

RQ: Whether there exists any sex differences in Norwegian students' perceptions of teacher unfairness using PISA 2015 data.

Methods

Sample

This study draws the Norwegian portion of student responses to the *PISA 2015 Student Questionnaire* (OECD, 2014) from *PISA 2015 Database* (OECD, 2018).

Out of a sample size of $N = 5456$, 428 entries contain missing values in some or all columns of interest, representing 7.84% of the dataset. Since the loss rate is greater than 5 percent, an investigation must be conducted over the pattern of missing data (Rubin, 1976). Little's MCAR Test (Little, 1988) produced strong evidence against the "missing completely at random" assumption ($\chi^2 = 174$, $p < 0.01$), suggesting the necessity of missing data treatment. Complement to this statistical evidence, there is behavioural concern that students who experienced or witnessed unfair teacher treatment may be more likely to skip this question than those who did not, either due to social desirability or out of fear of retribution.

This study followed Yang et al. (2012)'s advice in choosing multiple imputation as the missing data treatment. Students' maths, science and reading scores (10 plausible values each) are used in Stata's `mi impute ologit` command, with `W_FSTUWT` as sampling weights. Five plausible values are obtained from this procedure and the median outcome is chosen as the best approximation for each missing value.

Measures

Teacher unfairness appears as Question ST039 in the *Student questionnaire* (an extract is presented in Appendix A) and contains six sub-items (`ST039Q01NA` to `ST039Q06NA`). Students were instructed to report the frequency of various experiences related to teacher unfairness by select one response in a four-point Likert scale, ranging from "Never or

Table 2
Descriptive Statistics

Item	*n*			*M*			Median			*SD*			Skewness			Kurtosis		
Number	girl	boy	[%Δ]	girl	boy	[%Δ]	girl	boy	[%Δ]	girl	boy	[%Δ]	girl	boy	[%Δ]	girl	boy	[%Δ]
ST039Q01NA	2706	2750	[1%]	1.89	1.92	[2%]	2	2	[0%]	0.98	1.02	[4%]	0.79	0.75	[5%]	−0.52	−0.7	[30%]
ST039Q02NA	2706	2750	[1%]	1.88	2.01	[7%]	2	2	[0%]	0.89	0.97	[9%]	0.73	0.6	[19%]	−0.32	−0.68	[72%]
ST039Q03NA	2706	2750	[1%]	1.78	1.78	[0%]	1	1	[0%]	0.95	0.98	[3%]	0.98	1	[2%]	−0.15	−0.19	[25%]
ST039Q04NA	2706	2750	[1%]	1.38	1.68	[20%]	1	1	[0%]	0.76	0.99	[26%]	2.09	1.22	[55%]	3.58	0.21	[244%]
ST039Q05NA	2706	2750	[1%]	1.36	1.45	[6%]	1	1	[0%]	0.7	0.82	[16%]	2.05	1.83	[11%]	3.68	2.35	[44%]
ST039Q06NA	2706	2750	[1%]	1.39	1.45	[4%]	1	1	[0%]	0.73	0.83	[13%]	1.99	1.85	[7%]	3.35	2.46	[30%]

Note. This table is generated using the `describeBy( )` function in R's `psych` package. Percentage changes in square brackets [%Δ] are calculated by

$$\frac{\big||\text{Stat}_{\text{girl}}| - |\text{Stat}_{\text{boy}}|\big|}{|\text{Stat}_{\text{all}}|} \times 100\%.$$

Table 3
Correlation Table

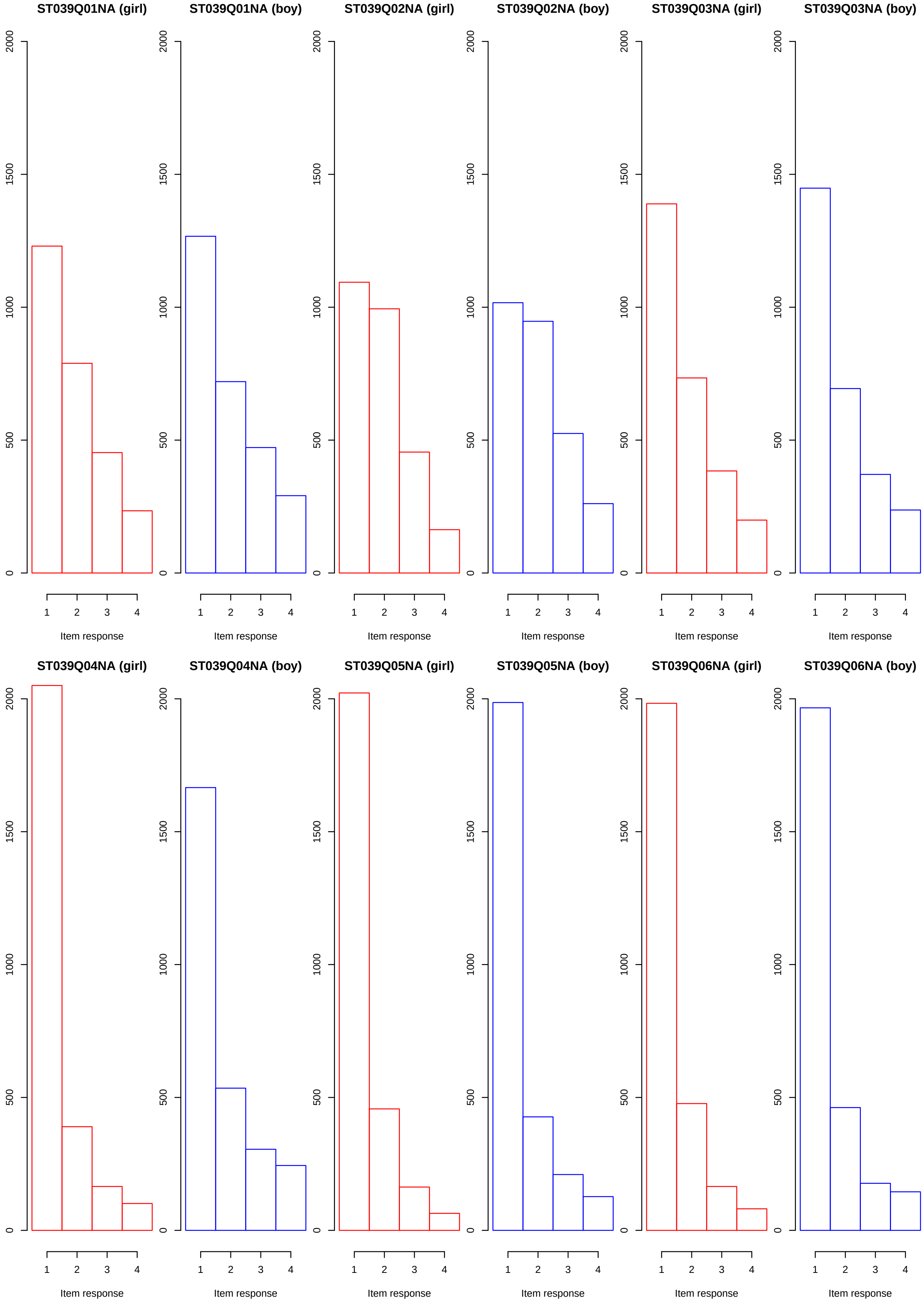
	ST039Q01NA	ST039Q02NA	ST039Q03NA	ST039Q04NA	ST039Q05NA	ST039Q06NA
ST039Q01NA	—					
ST039Q02NA	0.47 0.48 [2%]	—				
ST039Q03NA	0.46 0.43 [36%]	0.54 0.57 [5%]	—			
ST039Q04NA	0.34 0.43 [30%]	0.52 0.59 [13%]	0.53 0.60 [13%]	—		
ST039Q05NA	0.31 0.35 [12%]	0.39 0.43 [10%]	0.46 0.55 [18%]	0.53 0.53 [0%]	—	
ST039Q06NA	0.32 0.35 [9%]	0.40 0.47 [16%]	0.49 0.54 [10%]	0.53 0.60 [12%]	0.78 0.79 [1%]	—

Note. Item correlations are computed separately by sex ($\text{Corr}_{\text{girl}}$ Corr_{boy}). Percentage differences in square brackets are computed by

$$\frac{|\text{Corr}_{\text{girl}} - \text{Corr}_{\text{boy}}|}{\text{Corr}_{\text{all}}} \times 100\%.$$

Figure 1

Frequency Distribution (girl boy)



almost never” to “Once a week or more”. Students’ sex information is obtained from variable `ST004D01T`.

Statistical analyses

With the exception of the Stata (Version 16.1, StataCorp (2019)) multiple imputation procedure described above, this project mainly uses R (Version 3.6.3, R Core Team (2020)) for statistical analyses. The `pisa.select.merge( )` function in R package intsvy is used for country- and item-selective data import (Caro & Biecek, 2017). All IRT analyses are conducted using R’s mirt package (Chalmers, 2012).

This project employs generalised partial credit models (GPCM, Muraki (1992)) for all its IRT procedures. A GPCM is derived from a partial credit model (PCM, Masters (1982)), which uses the Rash model to specify the probability of success at the k-th step such that the item response function for $Y_i = j$ ($j = \{1, 2, 3, 4\}$) has the form

$$\mathbb{P}_{ij}(\theta) = \frac{\exp \sum_{k=1}^{j} (\theta - b_{ik})}{1 + \sum_{r=1}^{m} \left(\exp \sum_{k=1}^{r} (\theta - b_{ik}) \right)}. \tag{1}$$

A GPCM generalises the PCM by allowing the item-level discrimination parameter (a_i) to vary:

$$\mathbb{P}_{ij}(\theta) = \frac{\exp \sum_{k=1}^{j} [a_i(\theta - b_{ik})]}{1 + \sum_{r=1}^{m} \left(\exp \sum_{k=1}^{r} [a_i(\theta - b_{ik})] \right)}. \tag{2}$$

De Ayala (2009) recommends PCM over competing models due to its parsimonious nature. Since a relative small number of estimates are made per set of items, a PCM model can reach stable parameter estimations with sample sizes as small as 250 for three-category items (Choi et al., 1997). Additional statistical advantage of PCM over, for instance, graded response models (GRM, Samejima (1969)), is summarised by Masters and Wright (1997). By fixing the value of a_i to 1 across all items, a GPCM is reduced to a PCM. The GPCM is flexible by allowing the possibility of identifying item response options that may be redundant with each other.

Results

Descriptive Statistics and Correlations

Before conducting any IRT analyses, it is desirable to inspect the dataset. In tabular and graphical forms respectively, Table 2 and Figure 1 show the distribution of student responses, segregated by sex. All items display positive skewness, indicating teacher unfairness is uncommon in Norway. `ST039Q04NA` stands out as having a noticeable sex difference due to large difference in means, with boys reporting higher frequency of being disciplined more harshly by their teachers. Correlations in Table 3 highlights sizeable sex differences in `ST039Q01NA`–`ST039Q03NA`, and `ST039Q01NA`–`ST039Q04NA` pairs.

Model Fit and Item Fit Indices

Table 4 summarises the output of `M2( )` function as a model fit measure (Hansen et al., 2016). The p values of both sexes are less than 0.05, suggesting significant differences between observed and model-estimated data. Both models carry large RMSEAs (> 0.10) signaling poor fit at the model level. RMSEAs for item fit indices $S - X^2$, on the other hand, are small (Table 5), suggesting acceptable fit at the item level (Orlando & Thissen, 2000).

DIF

Table 6 summarises the main findings of the DIF investigation. `ST039Q03NA`, `ST039Q04NA` and `ST039Q06NA` have been identified as items with significant differences in response patterns between boys and girls (large χ^2 statistics, all $p < 0.01$). Category characteristic curves (CCC) in Figure 2 further highlight the different response patterns between boy and girls: In `ST039Q03NA`, the red CCCs appear left-shifted relative the blue CCCs while the reverse is true for `ST039Q04NA`. In `ST039Q06NA` the positions of the red CCCs are comparable to the blues’ but more “peaked”. These patterns suggest that while boys self-report to receive heavier discipline by their teachers, it is the impression of being thought of less smart that dominates girls’ teacher unfairness perceptions. Girls are also more likely than boys to interpret teacher comments as insulting. Statements about such sex differences can be made with high degrees of confidence because most CCC differentials appear over the intervals with the highest item information, evidenced by the plots of item information functions (IIF) in Figure 3.

Discussion

This study investigates sex differences in Norwegian students’ perceptions of teacher unfairness using response data from *2015 PISA Student Questionnaire*. With an IRT framework, DIF pattens have been identified for three out of the six items used by PISA to measure the construct of teacher unfairness. Dissimilar to prior research conducted in France (Lentillon et al., 2006), Germany and Israel (Resh & Dalbert, 2007), the current study does not find strong sex differences in Norwegian students’ perceptions over grades allocation. It does find evidence in partial support for OECD’s observation that boy are more likely than girls to report being treated unfairly by their teachers (OECD, 2017a)—although in the Norwegian context, such perceptions are limited to classroom discipline only. This study contributes to existing research by reporting female students’ tendency of feeling being considered less smart by their teachers and a higher propensity of perceiving teachers’ comments as insulting.

The DIF results of this study do not align neatly with taxonomies put forward by classroom justice literature. As summarised in Table 1, the presence of DIF does not strictly

Table 4

Model Fit Indices

	M2	df	p	RMSEA	RMSEA_5	RMSEA_95	SRMSR	TLI	CFI
girl	691	9	0	0.1674	0.1569	0.1781	0.0823	0.8806	0.9284
boy	490	9	0	0.1395	0.1291	0.1501	0.0684	0.9302	0.9581
all	1141	9	0	0.1518	0.1445	0.1593	0.0737	0.9106	0.9464

Note. This table is generated using mirt's `M2( )` function with option `type="C2"`.

Table 5

Item Fit Indices

Item	$S\text{-}X^2$		$df(S\text{-}X^2)$		RMSEA$(S\text{-}X^2)$		$p(S\text{-}X^2)$	
Number	girl	boy	girl	boy	girl	boy	girl	boy
ST039Q01NA	48.358	104.641	32	32	0.014	0.029	0.032	0.000
ST039Q02NA	75.849	80.980	29	28	0.024	0.026	0.000	0.000
ST039Q03NA	102.948	89.033	29	29	0.031	0.027	0.000	0.000
ST039Q04NA	126.004	141.732	30	29	0.034	0.038	0.000	0.000
ST039Q05NA	88.896	154.024	27	26	0.029	0.042	0.000	0.000
ST039Q06NA	117.812	146.334	29	27	0.034	0.040	0.000	0.000

Note. A `multipleGroup( )` function is used first to consolidate the original dataset by sex. This multiple group object is then received by mirt's `itemfit( )` function to generate the table above.

Table 6

Differential Item Functioning (DIF)

	AIC	AICc	SABIC	HQ	BIC	χ^2	df	p
ST039Q01NA	1.414	1.434	4.841	3.719	8.019	0.586	1	0.444
ST039Q02NA	0.728	0.748	4.155	3.033	7.332	1.272	1	0.259
ST039Q03NA	−14.860	−14.840	−11.433	−12.555	−8.256	16.860	1	0.000
ST039Q04NA	−89.877	−89.857	−86.450	−87.527	−83.272	91.877	1	0.000
ST039Q05NA	−0.764	−0.744	2.663	1.541	5.841	2.764	1	0.096
ST039Q06NA	−12.081	−12.061	−8.655	−9.777	−5.477	14.081	1	0.000

Note. This table is generated using mirt's `DIF( )` function. `ST039Q03NA`, `ST039Q04NA` and `ST039Q06NA` are identified as items possessing DIF based on large χ^2 statistics.

Figure 2

Category Characteristic Curves (girl boy all)

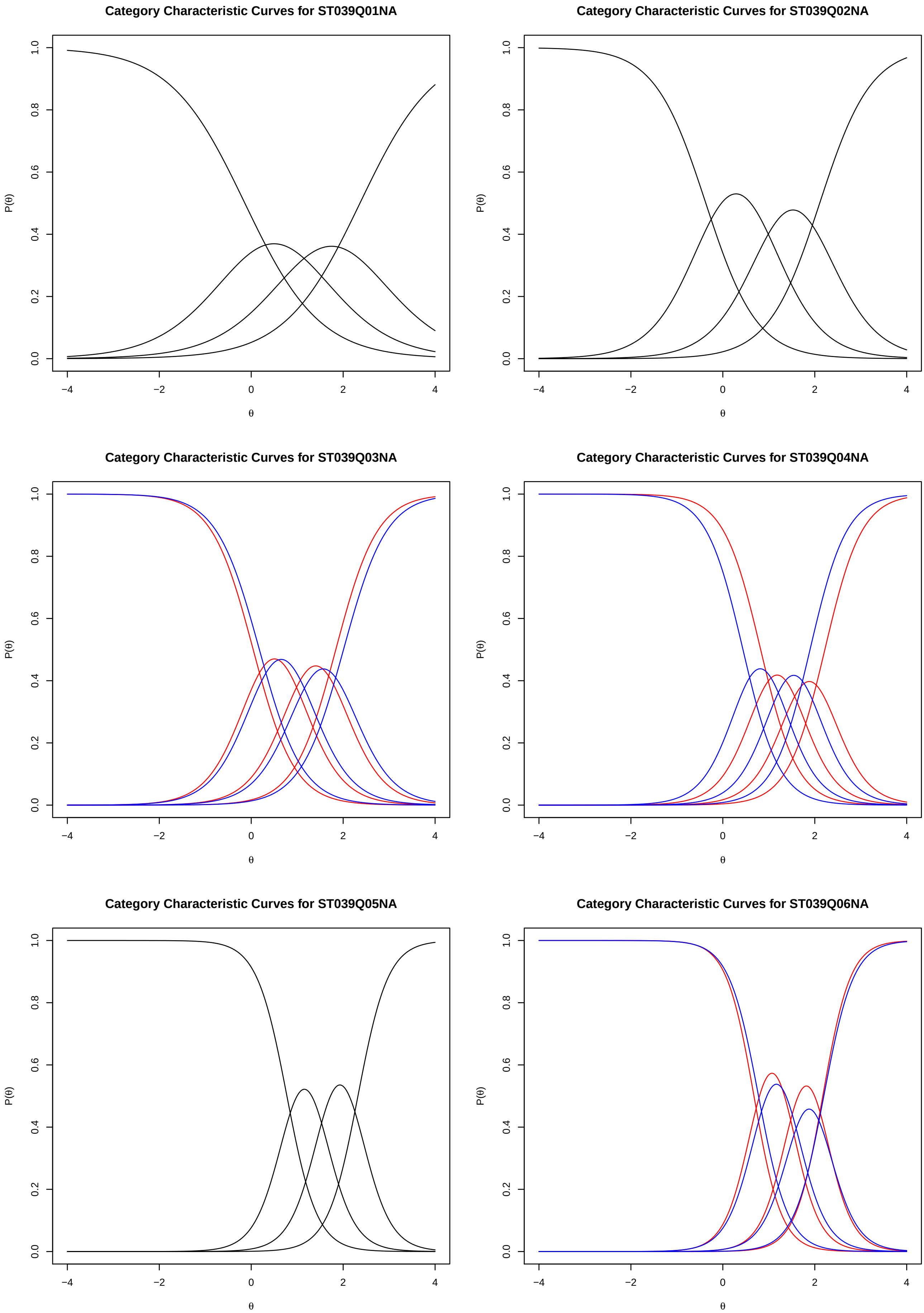
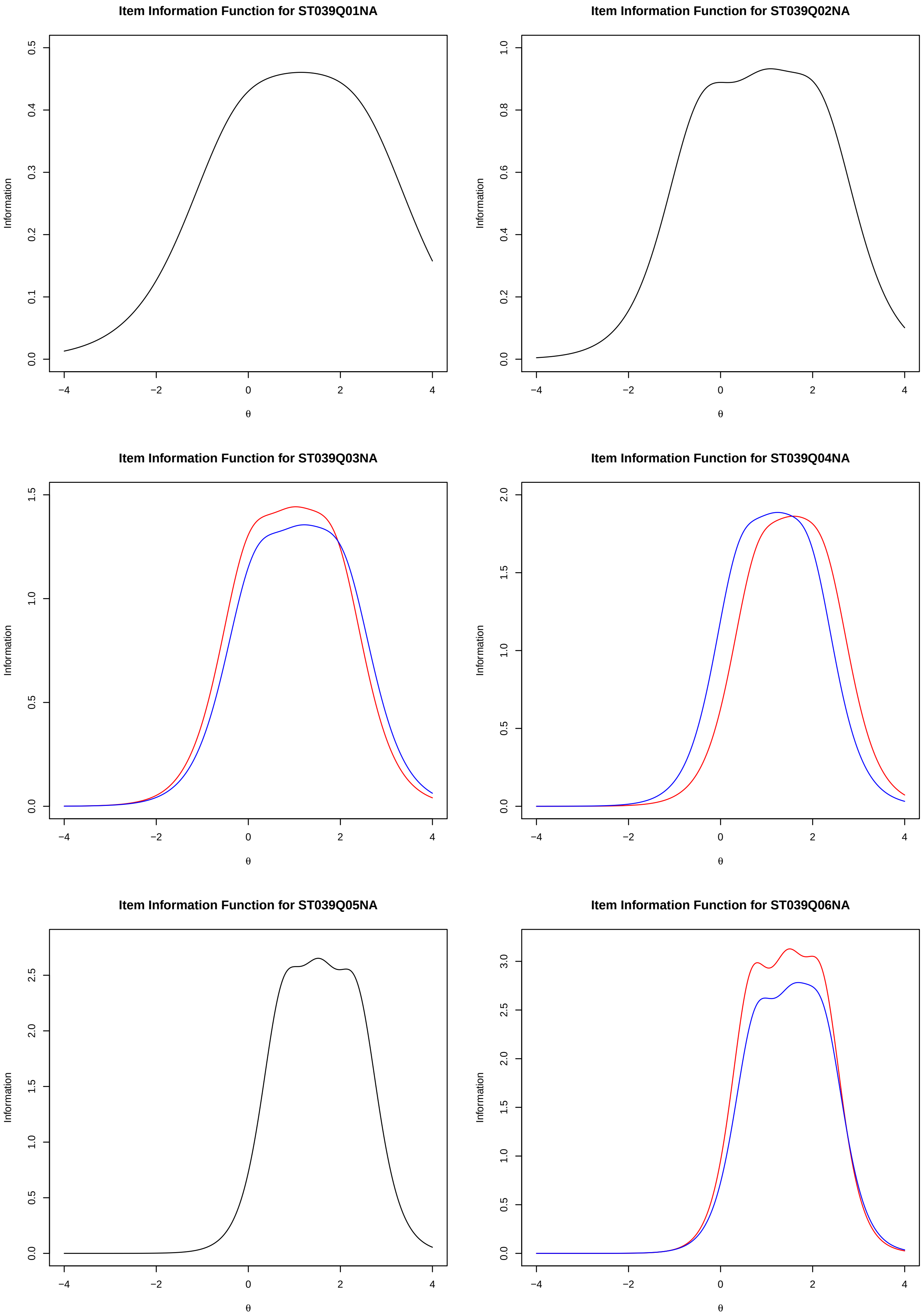


Figure 3

Item Information Functions (girl boy all)



follow either Chory-Assad and Paulsel (2004) or Horan et al. (2010)'s classification systems. It is interesting to observe, however, that little sex differences are reported over teacher actions that can be verified (e.g., teacher graded me harder); all three items with strong sex differences are products of students' internalisation (e.g., "I think that my teacher thinks that…"). One possible interpretation is that classroom justice is already strong in Norway; students' subjective assessment of teacher action is more likely to be the source of variation in teacher unfairness perceptions rather than actual teacher misconduct. In answering the research question "whether there exists any sex differences in Norwegian students' perceptions of teacher unfairness", the answer is "yes".

This project is also careful in carrying out the IRT methodology. Under the advice of Rubin (1976), an investigation has been conducted over the patterns of missing data. Multiple imputation is subsequently executed under the theoretical work of Yang et al. (2012) in order to address the non-random missing data problem. In addition, careful consideration is given for choosing GPCM as the main IRT model due to its parsimony, estimation efficiency, as well as flexibility. Item information is also taken into consideration when making statements about the presence of sex differences in student perceptions.

Limitations and Future Directions

This study is conservative in adopting a one-country data design to minimise variation introduced by cultural differences. The external validity of this research finding is therefore unclear—to what extent similar conclusions can be drawn to other countries, both similar to and different from Norway in social convention and cultural context. Future research may first of all compare student response data from other Nordic countries with Norway's, and then introduce countries further away to investigate whether there exist a gradient of perception patterns amongst world's students at large.

References

Angoff, W. H. (1993). Perspectives on differential item functioning methodology. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 3–23). Lawrence Erlbaum Associates.

Bock, R. D. & Zimowski, M. F. (1997). Multiple group IRT. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 433–448). Springer. https://doi.org/10.1007/978-1-4757-2691-6_25

Caro, D. H. & Biecek, P. (2017). intsvy: An R package for analyzing international large-scale assessment data. *Journal of Statistical Software*, *81*(7), 1–44. https://doi.org/10.18637/jss.v081.i07

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Chen, F. & Cui, Y. (2019). Investigating the relation of perceived teacher unfairness to science achievement by hierarchical linear modeling in 52 countries and economies. *Educational Psychology*, *40*(3), 273–295. https://doi.org/10.1080/01443410.2019.1652248

Choi, S. W., Cook, K. F. & Dodd, B. G. (1997). Parameter recovery for the partial credit model using MULTILOG. *Journal of Outcome Measurement*, *1*(2), 114–142. http://jampress.org/JOM_V1N2.pdf

Chory-Assad, R. M. (2002). Classroom justice: Perceptions of fairness as a predictor of student motivation, learning, and aggression. *Communication Quarterly*, *50*(1), 58–77. https://doi.org/10.1080/01463370209385646

Chory-Assad, R. M. & Paulsel, M. L. (2004). Antisocial classroom communication: Instructor influence and interactional justice as predictors of student aggression. *Communication Quarterly*, *52*(2), 98–114. https://doi.org/10.1080/01463370409370184

De Ayala, R. J. (2009). *The theory and practice of item response theory*. Builford Publications.

Hansen, M., Cai, L., Monroe, S. & Li, Z. (2016). Limited-information goodness-of-fit testing of diagnostic classification item response models. *British Journal of Mathematical and Statistical Psychology*, *69*(3), 225–252. https://doi.org/10.1111/bmsp.12074

Holland, P. W. & Wainer, H. (1993). *Differential item functioning*. Routledge.

Horan, S. M., Chory, R. M. & Goodboy, A. K. (2010). Understanding students' classroom justice experiences and responses. *Communication Education*, *59*(4), 453–474. https://doi.org/10.1080/03634523.2010.487282

Lentillon, V., Cogérino, G. & Kaestner, M. (2006). Injustice in physical education: Gender and the perception of deprivation in grades and teacher support. *Social Psychology of Education*, *9*(3), 321–339. https://doi.org/10.1007/s11218-005-5122-z

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. https://doi.org/10.1007/BF02296272

Masters, G. N. & Wright, B. D. (1997). The partial credit model. In W. J. van der Linden & R. K. Hambleton

(Eds.), *Handbook of modern item response theory* (pp. 101–121). Springer. https://doi.org/10.1007/978-1-4757-2691-6_6

Muraki, E. (1992). A generalized partial credit model: Application of and EM algorithm. *ETS Research Report Series*, *1992*(1), 1–30. https://doi.org/10.1002/j.2333-8504.1992.tb01436.x

OECD. (2014). *Student questionnaire for PISA 2015 (computer-based version) main survey version*. OECD Publishing. https://www.oecd.org/pisa/data/CY6_QST_MS_STQ_CBA_Final.pdf

OECD. (2017a). *PISA 2015 assessment and analytical framework: Science, reading, mathematic, financial literacy and collaborative problem solving* (Revised ed.). OECD Publishing. https://doi.org/10.1787/9724281820-en

OECD. (2017b). *PISA 2015 results (Volume III): Students' well-being*. OECD Publishing. https://doi.org/10.1787/9789264273856-en

OECD. (2018). *PISA 2015 database: Student questionnaire data file*. OECD Publishing. https://webfs.oecd.org/pisa/PUF_SPSS_COMBINED_CMB_STU_QQQ.zip

Orlando, M. & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, *24*(1), 50–64. https://doi.org/10.1177/01466216000241003

Peter, F. & Dalbert, C. (2010). Do my teachers treat me justly? Implications of students' justice experience for class climate experience. *Contemporary Educational Psychology*, *35*(4), 297–305. https://doi.org/10.1016/j.cedpsych.2010.06.001

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Resh, N. & Dalbert, C. (2007). Gender differences in sense of justice about grades: A comparative study of high school students in Israel and Germany. *Teachers College Record*, *109*(2), 322–432. https://www.researchgate.net/publication/286356463

Roussos, L. A. & Stout, W. F. (1996). Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel-Haenszel Type I Error performance. *Journal of Educational Measurement*, *33*(2), 215–230. https://doi.org/10.1111/j.1745-3984.1996.tb00490.x

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.2307/2335739

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, *34*(S1), 1–97. https://doi.org/10.1007/BF03372160

Shadish, W. R., Cook, T. D. & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference* (2nd Ed.). Cengage Learning.

StataCorp. (2019). *Stata statistical software: Release 16*. College Station, TX, StataCorp LLC.

Thissen, D., Steinberg, L. & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 67–113). Lawrence Erlbaum Associates.

Walstad, W. B. & Rebeck, K. (2017). The test of financial literacy: Development and measurement characteristics. *The Journal of Economic Education*, *48*(2), 113–122. https://doi.org/10.1080/00220485.2017.1285739

Yang, J. S., Hansen, M. & Cai, L. (2012). Characterizing sources of uncertainty in item response theory scale scores. *Educational and Psychological Measurement*, *72*(2), 264–290. https://doi.org/10.1177/0013164411410056

Appendix
Teacher Unfairness (Question ST039)

ST039 During the past 12 months, how often did you have the following experiences at school?

(Please select one response in each row.)

		Never or almost never	*A few times a year*	*A few times a month*	*Once a week or more*
ST039Q01NA	Teachers called on me less often than they called on other students.	☐1	☐2	☐3	☐4
ST039Q02NA	Teachers graded me harder than they graded other students.	☐1	☐2	☐3	☐4
ST039Q03NA	Teachers gave me the impression that they think I am less smart than I really am.	☐1	☐2	☐3	☐4
ST039Q04NA	Teachers disciplined me more harshly than other students.	☐1	☐2	☐3	☐4
ST039Q05NA	Teachers ridiculed me in front of others.	☐1	☐2	☐3	☐4
ST039Q06NA	Teachers said something insulting to me in front of others.	☐1	☐2	☐3	☐4

Appendix D

Selected Work: Multilevel Modelling (MAE4112) Assignment

1 Conceptual problems

1. Multilevel modelling terms and conditions

   Explain the following terms and conditions.

   (a) Cross-level interaction effect

   Solution: Consider a two-level model containing both $L1$ and $L2$ predictors, X and W, respectively:

   $$\begin{cases} L2: \begin{pmatrix}\beta_{0j}\\ \beta_{1j}\end{pmatrix} = \begin{pmatrix}\gamma_{00} & \gamma_{01}\\ \gamma_{10} & \gamma_{11}\end{pmatrix}\begin{pmatrix}1\\ W_j\end{pmatrix} + \begin{pmatrix}u_{0j}\\ u_{1j}\end{pmatrix} \\ L1: \ Y_{ij} = \begin{pmatrix}1 & X_{ij}\end{pmatrix}\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix} + r_{ij} \end{cases}.$$

   A cross-level interaction term investigates whether the $L2$ predictor W would have any effect on the nature or strength of the relationship between $L1$ predictor X and $L1$ outcome variable Y.

   Out of the fixed effect (FE) matrix $\mathbf{\Gamma}$:

   $$\mathbf{\Gamma} = \begin{pmatrix}\gamma_{00} & \gamma_{01}\\ \gamma_{10} & \gamma_{11}\end{pmatrix} = \begin{pmatrix}\text{“true mean”} & \text{FE for } L2\ W\\ \text{FE for } L1\ X & \text{FE for } X\times W\end{pmatrix},$$

   it is the coefficient γ_{11} that we are interested in as a measure for the cross-level interaction effect.

   (b) Contextual effect

   Solution: A higher level (group) variable is said to have a contextual effect when it “trickles down” onto the lower level (individual) and affect its outcome, even after having lower level confounders controlled for.

   Variables showing the contextual effect can take one of two forms: an *integral* group-level variable can be thought as a “true $L2$ variable”, such as school size, that has little meaning at individual student-level; a *derived* group-level variable on the other hand is “made up” by aggregating an $L1$ variable and plug its mean into the $L2$ equation — for example, students’ individual socio-economic status can be averaged into a school’s average SES score, this school-level SES variable is a derived group-level variable.

   (c) Group mean-centered $L1$ predictor

   Solution: At $L1$, instead of using predictor X_{ij} at its original level, we may sometimes wish to “re-position” it about its mean. In order to construct a group mean-centered version, we first of all sort X_{ij} by their group marker j and compute the average $\overline{X}_{\cdot j}$ separately for each group. We then subtract these group means from the corresponding data to obtain $X_{ij} - \overline{X}_{\cdot j}$.

   When studying the effect of poverty, for example, the researcher may be more interested in people’s financial positions *relative to their peers* rather than in their absolute terms. When the research questions are centred on differences or deviations from some local norms, group mean-centered variables often provide better intuition.

(d) Multivariate normal distribution of the $L2$ random effects

Solution: In a two-level model, the outcome variable y can be decomposed into a fixed effect, $L2$ random effects and an $L1$ random effect:

$$\begin{cases} L2: & \boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{z} + \boldsymbol{u} \\ L1: & y = \boldsymbol{x}^\intercal \boldsymbol{\beta} + r \end{cases},$$

with total equation:

$$y = \underbrace{\boldsymbol{x}^\intercal \boldsymbol{\Gamma}\boldsymbol{z}}_{\text{fixed}} + \underbrace{\boldsymbol{x}^\intercal \boldsymbol{u}}_{L2 \text{ random}} + \underbrace{r}_{L1 \text{ random}}.$$

At $L2$:

$$\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{z} + \boldsymbol{u}$$

$$\underset{(k+1)\times 1}{\overset{\text{parameters}}{\begin{pmatrix} \beta_{0j} \\ \beta_{1j} \\ \vdots \\ \beta_{kj} \end{pmatrix}}} = \underset{(k+1)\times(k+1)}{\overset{\text{fixed}}{\begin{pmatrix} \gamma_{00} & \gamma_{01} & \cdots & \gamma_{0k} \\ \gamma_{10} & \gamma_{11} & \cdots & \gamma_{1k} \\ & \vdots & & \\ \gamma_{k0} & \gamma_{k1} & \cdots & \gamma_{kk} \end{pmatrix}}} \underset{(k+1)\times 1}{\overset{\text{data}}{\begin{pmatrix} 1 \\ z_{1j} \\ \vdots \\ z_{kj} \end{pmatrix}}} + \underset{(k+1)\times 1}{\overset{\text{random}}{\begin{pmatrix} u_{0j} \\ u_{1j} \\ \vdots \\ u_{kj} \end{pmatrix}}},$$

whose random effects $\boldsymbol{u}$ are assumed to follow a multivariate normal distribution with zero means ($\boldsymbol{\mu} = \boldsymbol{0}$) and a finite variance-covariance matrix $\boldsymbol{\Sigma}$:

$$\boldsymbol{u} = \begin{bmatrix} u_{0j} \\ u_{1j} \\ \vdots \\ u_{kj} \end{bmatrix} \sim \text{MVN} \left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u_0}^2 & \sigma_{u_0 u_1} & \cdots & \sigma_{u_0 u_k} \\ \sigma_{u_0 u_1} & \sigma_{u_1}^2 & \cdots & \sigma_{u_1 u_k} \\ & \vdots & & \\ \sigma_{u_0 u_k} & \sigma_{u_1 u_k} & \cdots & \sigma_{u_k}^2 \end{bmatrix} \right).$$

It is important to impose such assumptions on $L2$ errors ($\mathbb{E}[\boldsymbol{u}] = \boldsymbol{0}$) and $L1$ error ($\mathbb{E}[r] = 0$) because it is otherwise infeasible to estimate the fixed effect coefficients $\widehat{\boldsymbol{\Gamma}}$:

$$\mathbb{E}[\boldsymbol{\Gamma}] = \overbrace{\left[\left(\boldsymbol{x}\boldsymbol{x}^\intercal\right)^{-1} \boldsymbol{x} \right] y \left[\boldsymbol{z}^\intercal \left(\boldsymbol{z}\boldsymbol{z}^\intercal\right)^{-1} \right]}^{\text{data}} - \underbrace{\mathbb{E}[\boldsymbol{u}]}_{=\boldsymbol{0}} \left[\boldsymbol{z}^\intercal \left(\boldsymbol{z}\boldsymbol{z}^\intercal\right)^{-1} \right] - \left[\left(\boldsymbol{x}\boldsymbol{x}^\intercal\right)^{-1} \boldsymbol{x} \right] \underbrace{\mathbb{E}[r]}_{=0} \left[\boldsymbol{z}^\intercal \left(\boldsymbol{z}\boldsymbol{z}^\intercal\right)^{-1} \right].$$

(e) Reliability of aggregated variables ICC_2

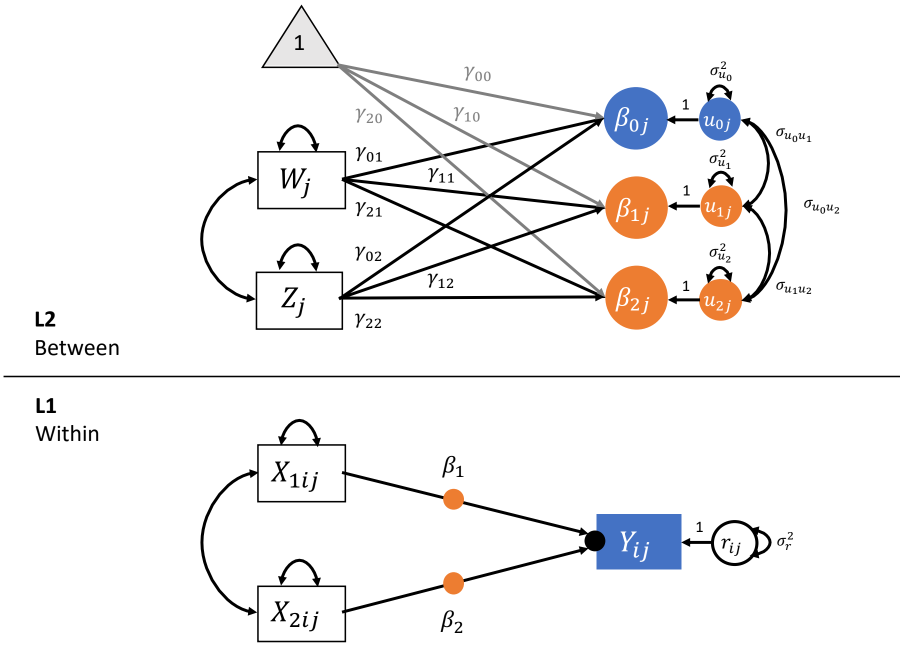
Solution: When forming group (e.g., schools by `SCHOOLID`) means, we wish to know how well these means are in distinguishing schools. We construct a *group mean reliability* metric ICC_2 by first of all fitting a null model to the variable of interest (e.g., `SES`): `SES ~ 1 + (1 | SCHOOLID)`, and compute ICC_2 via

$$\text{ICC}_{2j} = \lambda_j = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_r^2 / n_j},$$

where n_j is the group size.
A low ICC_2 indicates potential bias in $L2$ estimations.

2. Path diagram representation of multilevel models

The following path diagram represents a multilevel model with two levels ($L1$ and $L2$):



(a) Formulate the $L1$ and $L2$ model equations, including the assumptions on the distributions of the $L1$ and $L2$ residuals.

Solution:

$L1$**:**

$$\begin{aligned} Y_{ij} &= \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + r_{ij} \\ &= \begin{pmatrix} 1 & X_{1ij} & X_{2ij} \end{pmatrix} \begin{pmatrix} \beta_{0j} \\ \beta_{1j} \\ \beta_{2j} \end{pmatrix} + r_{ij}, \end{aligned}$$

with

$$r_{ij} \sim \mathcal{N}(0, \sigma_r^2).$$

$L2$**:**

$$\begin{pmatrix} \beta_{0j} \\ \beta_{1j} \\ \beta_{2j} \end{pmatrix} = \begin{pmatrix} \gamma_{00} & \gamma_{01} & \gamma_{02} \\ \gamma_{10} & \gamma_{11} & \gamma_{12} \\ \gamma_{20} & \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} 1 \\ W_j \\ Z_j \end{pmatrix} + \begin{pmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{pmatrix},$$

with

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{pmatrix} \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{u_0} & \sigma_{u_0u_1} & \sigma_{u_0u_2} \\ \sigma_{u_0u_1} & \sigma^2_{u_1} & \sigma_{u_1u_2} \\ \sigma_{u_0u_2} & \sigma_{u_1u_2} & \sigma^2_{u_2} \end{bmatrix} \right).$$

(b) How many random and fixed effects are estimated in this model?

Solution: Observe the total equation

$$Y_{ij} = \underset{1\times 3}{\overset{\text{data}}{\begin{pmatrix}1 & X_{1ij} & X_{2ij}\end{pmatrix}}} \underset{3\times 3}{\overset{\text{fixed}}{\begin{pmatrix}\gamma_{00} & \gamma_{10} & \gamma_{20} \\ \gamma_{01} & \gamma_{11} & \gamma_{21} \\ \gamma_{02} & \gamma_{12} & \gamma_{22}\end{pmatrix}}} \underset{3\times 1}{\overset{\text{data}}{\begin{pmatrix}1 \\ W_j \\ Z_j\end{pmatrix}}} + \underset{1\times 3}{\overset{\text{data}}{\begin{pmatrix}1 & X_{1ij} & X_{2ij}\end{pmatrix}}} \underset{3\times 1}{\overset{\text{random}}{\begin{pmatrix}u_{0j} \\ u_{1j} \\ u_{2j}\end{pmatrix}}} + \underset{1\times 1}{\overset{\text{random}}{r_{ij}}}$$

1. Number of random effects: $1 \times 3 + 1 \times 1 = 4$,
2. Number of fixed effects: $3 \times 3 = 9$.

(c) Formulate a research question researchers could address with this model. Please define the levels of analysis ($L1$ and $L2$) and what the variables in the model may represent.

Solution: Imagine I were a tax researcher working for the government, wishing to develop a fraud detection screening tool. Since firms are grouped into industries — while it is generally difficult to compare firms from different industries (Norwegian Air vs Novartis), those within the same industry should have relatively comparable account structure, revenue stream and expenditure patterns (e.g., Norwegian vs SAS) — this intra- vs inter-industry differences make multilevel modelling a desirable starting point in developing the detection tool.

At the industry level ($L2$), I hypothesise that the sizes of fixed assets makes a systematic difference in tax calculation because industries with large property, plant and equipment (`PPE`) accounts (e.g., airline industry) would have large depreciation expenditure. Similarly, industries with large research and development (`R&D`) activities (e.g., pharmaceutical companies) would have different tax treatment for their `R&D` outgoings.

At the firm level ($L1$), I propose expenditure on wage and salaries (`wage`) as a good proxy for firm size. (I once considered using number of employees as a measure but, as politically incorrect as it sounds, not all employees contribute to the revenue figures equally; therefore `wage` can be thought as a "weighted sum" of employees.) Another firm-level variable that attracts tax authority's attention would be marketing (`mktg`) expenditure as a measure of firm's maturity (young firms advertise more to seek market share).

My research interest would be in r_{ij}, that is, having controlled for both industry-level variations, firm size and stages of growth, large deviation from benchmarks often warrants further investigation. After audit, I may evaluate my original research question: "Would multilevel residuals provide effective early warnings for tax fraud?"

2 Applied problems

1. Grade retention and mathematics achievement

   The U.S. Sustaining Effects Study examined the relation between students' grade retention and their achievement in mathematics for large and representative samples. The data (`egsingle`) contain the following variables:

   - `schoolid`: Identifier of the school
   - `childid`: Identifier of the student
   - `math`: Mathematics achievement score (based on an IRT scale)
   - `retained`: Binary variable indicating whether or not the student has retained in a grade (1 = retention, 0 = no retention)
   - `size`: Number of students in the school
   - `lowinc`: Percentage of students with a low-income background in the school

   (a) Determine the level at which these variables were measured.

   Solution:

Variable	Student level	School level
`schoolid`		✓
`childid`	✓	
`math`	✓	
`retained`	✓	
`size`		✓
`lowinc`		✓

   Analysing a subset of the data ($N = 7230$ students in 60 schools), a researcher specified the following multilevel model (`MLM2`) using lme4:

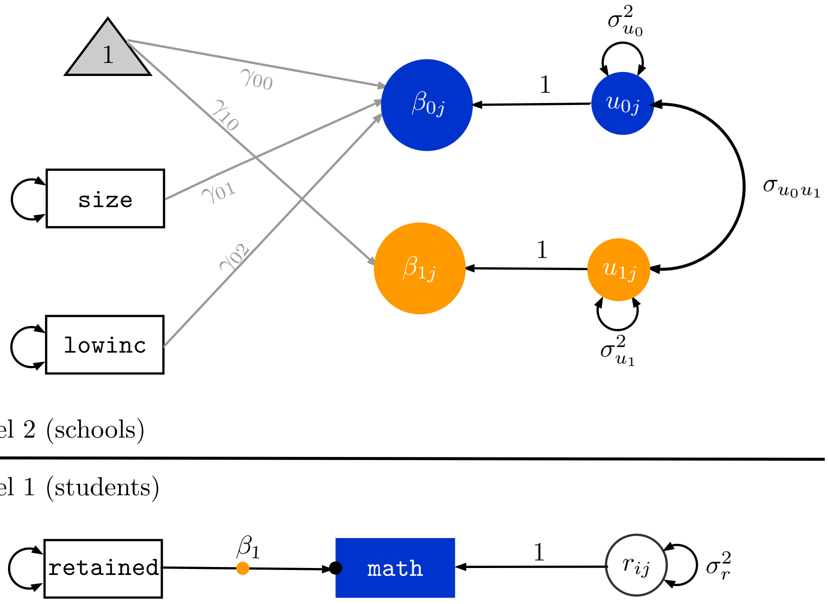
```r
# Model specification
MLM2 <- lmer(math ~ 1 + retained + size + lowinc + (1 + retained | schoolid), data=
    egsingle)

# Summarise the results
summary(MLM2)
```

   (b) Formulate the $L1$ and $L2$ model equations underlying this model.

   Solution:

$$\begin{cases} L2: \beta_{0j} = \gamma_{00} + \gamma_{01}\texttt{size}_{ij} + \gamma_{02}\texttt{lowinc}_{ij} + u_{0j} \\ \qquad \beta_{1j} = \gamma_{10} + u_{ij} \\ L1: \texttt{math}_{ij} = \beta_{0j} + \beta_{1j}\texttt{retained}_{ij} + r_{ij} \end{cases}$$

(c) Draw the corresponding path diagram.



(d) Estimating the model (`MLM2`) using lme4, the researchers obtained the following output:

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: math ~ 1 + retained + size + lowinc + (1 + retained | schoolid)
   Data: egsingle

REML criterion at convergence: 25760.6

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1248 -0.6945 -0.0246  0.7092  3.8164

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 schoolid (Intercept) 0.2164   0.4651
          retained1   0.1574   0.3967   -0.80
 Residual             2.0136   1.4190
Number of obs: 7230, groups:  schoolid, 60

Fixed effects:
              Estimate Std. Error         df t value Pr(>|t|)
(Intercept)  3.589e-01  1.904e-01  5.686e+01   1.885   0.0645 .
retained1   -1.185e+00  9.765e-02  4.468e+01 -12.140 9.55e-16 ***
size         1.209e-05  1.844e-04  5.329e+01   0.066   0.9480
lowinc      -1.104e-02  2.184e-03  5.715e+01  -5.053 4.76e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
          (Intr) retnd1 size
retained1 -0.014
size      -0.452 -0.050
lowinc    -0.712 -0.130 -0.213
```

Which of the following conclusions can be drawn/are true?

1. Students who retained in a grade scored significantly lower on the math achievement test (`math`) than those who did not retain, after controlling for school size (`size`) and the proportion of low-income students in the school (`lowinc`).
2. Neither the proportion of low-income students (`lowinc`) nor the size of the school (`size`) were significantly related to the math achievement test score (`math`) between schools after controlling for retention (`retained`).
3. School size (`size`) explains significant variation in the random slopes (ie, the relation between `math` and `retained`).

Solution: Statement 1. is correct.

(e) In a second run of the analyses, the researchers modified the model (`MLM2`) slightly into model (`MLM1`) as follows:

```
# Model specification
MLM1 <- lmer(math ~ 1 + retained + size + lowinc + (1 | schoolid), data=egsingle)

# Summarise the results
summary(MLM1)
```

Describe the difference between models `MLM1` and `MLM2` conceptually.

Solution: `MLM1` and `MLM2` differ mainly in the variable `retained`. `MLM1` is a random intercept model while `MLM2` allows `retained` to vary both in intercepts and in slopes.

(f) To determine which of these two models (`MLM1` vs `MLM2`) represents the data better, the researchers performed a likelihood-ratio test and obtained the following results:

```
# Model comparison
anova(MLM2, MLM1)
```

```
Data: egsingle
Models:
MLM1: math ~ 1 + retained + size + lowinc + (1 | schoolid)
MLM2: math ~ 1 + retained + size + lowinc + (1 + retained | schoolid)
     Df   AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
MLM1  6 25756 25797 -12872    25744
MLM2  8 25744 25799 -12864    25728 15.768      2  0.0003768 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Decide on which of these two models represents the data better.

Solution: `MLM2` has smaller AIC, log-likelihood and deviance — all indicators suggesting better fit. Its BIC is marginally higher than that of `MLM1` but this could simply be an unfortunate result from how BIC is formulated: it provides "double penalty" for both parameters (q) and sample size (N) while AIC only penalises q. Since there is no reason to suggest sample size should be a cause for concern in this study, I prefer AIC to BIC as a model selection criterion. Finally, $\chi^2_2 = 15.768$ also suggests `MLM1` and `MLM2` are not equally good at fitting the data. I therefore conclude that `MLM2` represents a better model given this particular set of data.

(g) Comparing model `MLM1` with the null model of `math` (`Null`), the researcher wanted to compute the variance explanation at $L1$ (ie, the student level). They obtained the following model outputs:

Model `MLM1`:

```
# Model specification
MLM1 <- lmer(math ~ 1 + retained + size + lowinc + (1 | schoolid), data=egsingle)

# Summarize the results
summary(MLM1)
```

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: math ~ 1 + retained + size + lowinc + (1 | schoolid)
   Data: egsingle

REML criterion at convergence: 25776.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1156 -0.6945 -0.0279  0.7152  3.8186

Random effects:
 Groups   Name        Variance Std.Dev.
 schoolid (Intercept) 0.1751   0.4184
 Residual             2.0229   1.4223
Number of obs: 7230, groups:  schoolid, 60

Fixed effects:
              Estimate Std. Error         df t value Pr(>|t|)
(Intercept)  4.447e-01  1.892e-01  5.190e+01   2.351   0.0226 *
retained1   -1.059e+00  7.749e-02  7.220e+03 -13.664  < 2e-16 ***
size         1.204e-04  1.890e-04  5.220e+01   0.637   0.5268
lowinc      -1.305e-02  2.189e-03  5.151e+01  -5.963 2.27e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
          (Intr) retnd1 size
retained1 -0.011
size      -0.462  0.017
lowinc    -0.707 -0.027 -0.225
```

Null model:

```
# Model specification
Null <- lmer(math ~ 1 + (1 | schoolid), data=egsingle)

# Summarise the results
summary(Null)
```

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: math ~ 1 + (1 | schoolid)
```

```
   Data: egsingle

REML criterion at convergence: 25958.8

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.0441 -0.7154 -0.0313  0.7125  3.7441

Random effects:
 Groups   Name        Variance Std.Dev.
 schoolid (Intercept) 0.353    0.5941
 Residual             2.072    1.4393
Number of obs: 7230, groups:  schoolid, 60

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept) -0.49835    0.07948 55.29993   -6.27 5.75e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Compute the $L1$ variance explanations using the Bryk and Raudenbush (1992) and the Snijders and Bosker (2012) formulas respectively.

Solution:

Bryk and Raudenbush (1992), R^2_{L1}

$$\begin{aligned} R_1^2 &= \frac{\sigma_r^2(\text{null}) - \sigma_r^2(\text{RI})}{\sigma_r^2(\text{null})} \\ &= \frac{2.072 - 2.0229}{2.072} \\ &= 0.0237 \end{aligned}$$

Snijders and Bosker (2012), R^2_{L1}

$$\begin{aligned} R_1^2 &= 1 - \frac{\sigma_r^2(\text{RI}) + \sigma_{u0}^2(\text{RI})}{\sigma_r^2(\text{null}) + \sigma_{u0}^2(\text{null})} \\ &= 1 - \frac{2.0229 + 0.1751}{2.072 + 0.353} \\ &= 0.0936 \end{aligned}$$

2. Taut et al. (2019) conducted a study of the relation between student achievement and the competence of the teachers in classrooms. The overall sample contained $N = 1124$ Chilean students (i) in 48 classrooms (j).

   Taut et al. (2019) included three variables in their analyses:

   - $\texttt{SEPA\_Pre}_{ij}$: Standardised test score in Mathematics and Spanish (pre-test)
   - $\texttt{SEPA\_Post}_{ij}$: Standardised test score in Mathematics and Spanish (post-test)
   - $\texttt{TeacherCat}_{j}$: Teacher competence coded as 0 (basic) and 1 (competent) for each classroom

   The authors estimated three multilevel regression models:

   Model 1

   $$\texttt{SEPA\_Post}_{ij} = \beta_{0j} + \beta_1 \cdot \texttt{SEPA\_Pre}_{ij} + \varepsilon_{ij}$$
   $$\beta_{0j} = \gamma_{00} + \gamma_{01} \cdot \overline{\texttt{SEPA\_Pre}}_{j} + u_{0j}$$

   where

   $$u_{0j} \sim \mathcal{N}(0, \sigma^2_{u0}) \text{ and } \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2_{\varepsilon})$$

   Model 2

   $$\texttt{SEPA\_Post}_{ij} = \beta_{0j} + \beta_1 \cdot \texttt{SEPA\_Pre}_{ij} + \varepsilon_{ij}$$
   $$\beta_{0j} = \gamma_{00} + \gamma_{01} \cdot \overline{\texttt{SEPA\_Pre}}_{j} + \gamma_{02} \cdot \texttt{TeacherCat}_{j} + u_{0j}$$

   where

   $$u_{0j} \sim \mathcal{N}(0, \sigma^2_{u0}) \text{ and } \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2_{\varepsilon})$$

   Model 3

   $$\texttt{SEPA\_Post}_{ij} = \beta_{0j} + \beta_1 \cdot \texttt{SEPA\_Pre}_{ij} + \varepsilon_{ij}$$
   $$\beta_{0j} = \gamma_{00} + \gamma_{01} \cdot \overline{\texttt{SEPA\_Pre}}_{j} + \gamma_{02} \cdot \texttt{TeacherCat}_{j} + u_{0j}$$
   $$\beta_{1j} = \gamma_{10} + u_{1j}$$

   where

   $$u_{0j} \sim \mathcal{N}(0, \sigma^2_{u0}),\; u_{1j} \sim \mathcal{N}(0, \sigma^2_{u1}),\; \operatorname{cov}(u_{0j}, u_{1j}) = 0, \;\text{ and } \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2_{\varepsilon})$$

   (a) Assign all variables in thee models to either the student or the classroom level.

Variable	Student level	School level
$\texttt{SEPA\_Post}_{ij}$	✓	
$\overline{\texttt{SEPA\_Pre}}_{j}$		✓
$\texttt{TeacherCat}_{j}$		✓
$\texttt{SEPA\_Pre}_{ij}$	✓	

   (b) Which of the following statements about these models are true?

   1. Model 1 assumes that the intercept of SEPA_Post varies between classrooms.
   2. In all models, the relation between SEPA_Post and TeacherCat is fixed (ie, constant).
   3. The random effects in Model 3 are assumed to be independent.

4. Model 1 represents a contextual model.
5. In Model 2, the coefficient γ_{02} indicates the difference between basic and competent teachers in the average classroom `SEPA_Post` score after controlling for the average classroom `SEPA_Pre` score.
6. In Model 3, the coefficient γ_{10} represents the average classroom `SEPA_Post` score.

Solution: Statements 3., 4. and 5. are true.

3. Create a task for an exam on multilevel regression models. Formulate the task, provide its solution, and propose a scoring. Which skills does your task require?

Exam task

Examine the following R output and complete the table below:

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: mAch ~ 1 + cses * sector + (1 | school)
   Data: Hsb82

REML criterion at convergence: 46650.5

Scaled residuals:
     Min       1Q   Median       3Q      Max
-3.04661 -0.72723  0.01725  0.75624  3.00331

Random effects:
 Groups   Name        Variance Std.Dev.
 school   (Intercept)  6.742   2.597
 Residual             36.819   6.068
Number of obs: 7185, groups:  school, 160

Fixed effects:
                     Estimate Std. Error        df t value Pr(>|t|)
(Intercept)           11.3930     0.2929  158.4641  38.897  < 2e-16 ***
cses                   2.7821     0.1446 7020.0641  19.242  < 2e-16 ***
sectorCatholic         2.8053     0.4394  153.7145   6.385 1.93e-09 ***
cses:sectorCatholic   -1.3486     0.2184 7020.0641  -6.174 7.02e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) cses   sctrCt
cses         0.000
sectorCthlc -0.667  0.000
css:sctrCth  0.000 -0.662  0.000
```

Parameter	Estimate (two decimal points)
γ_{00}	
γ_{01}	
γ_{10}	
γ_{11}	
σ_r^2	
$\sigma_{u_0}^2$	

Solution and scoring

Parameter	Estimate (two decimal points)
γ_{00}	11.39
γ_{01}	2.81
γ_{10}	2.78
γ_{11}	-1.35
σ_r^2	36.82
$\sigma_{u_0}^2$	6.74

One mark each. Half mark if not following "two decimal points" instruction.

Required skills to solve the task

This exam task specifically tests learner's ability to extract model information from R output.

In order to distinguish true understanding from memorisation, I swapped positions of $L1$ FE γ_{10} with $L2$ FE γ_{01} and of $L1$ error variance $\sigma_{u_0}^2$ with $L2$ error variance σ_r^2.

I expect the marks to follow a gradient based on the level of true understanding:

Item	By memory	Some understanding	Full understanding
γ_{00}	✓	✓	✓
γ_{01}			✓
γ_{10}			✓
γ_{11}	✓	✓	✓
σ_r^2		✓	✓
$\sigma_{u_0}^2$		✓	✓
Mark	2/6	4/6	6/6
Percentage	33%	67%	100%

3 Data-analytic problem

1. High School and Beyond Study 1982

   The High School and Beyond Study was conducted amongst $N = 7185$ students from 160 schools in 1982. The following variables were measured:

   - `school`: Identifier of the school
   - `minrty`: Binary variable indicating whether students belong to a minority (`Yes`, `No`)
   - `sx`: Binary variable indicating students' gender (`Male`, `Female`)
   - `ses`: Index of students' socio-economic status
   - `mAch`: Students' mathematics achievement test score
   - `meanses`: School-average socio-economic status (ie, the group mean of `ses`)
   - `sector`: Binary variable indicating the type of school (`Public`, `Catholic`)
   - `cses`: Group mean-centred `ses` variable (ie, `cses` = `sees` − `meanses`)

   The full dataset is stored in the R packcage mlmRe under the name Hsb82. You may call the data using the following R code:

```
# Load the R package
library(mlmRev)

# Read and attach the data
data(Hsb82); attach(Hsb82)
```

   In the following tasks, make sure to use the group mean-centred SES variable (`cses`).

   Contextual effect of `cses`:

   Test the hypothesis that the school-average socio-economic status (`meanses`) has a contextual effect on students' individual mathematics achievement (`mAch`).

   (a) Provide the R code of the model you have used to test this hypothesis in lme4:

   Solution:

```
contextual <- lmer(mAch ~ 1 + cses + meanses + (1 | school), data=Hsb82)
summary(contextual)
```

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: mAch ~ 1 + cses + meanses + (1 | school)
   Data: Hsb82

REML criterion at convergence: 46568.6

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1666 -0.7254  0.0174  0.7558  2.9454

Random effects:
 Groups   Name        Variance Std.Dev.
 school   (Intercept)  2.693   1.641
```

```
 Residual             37.019   6.084
Number of obs: 7185, groups:  school, 160

Fixed effects:
             Estimate Std. Error        df t value Pr(>|t|)
(Intercept)   12.6833     0.1494  153.6518   84.91   <2e-16 ***
cses           2.1912     0.1087 7021.5092   20.16   <2e-16 ***
meanses        5.8662     0.3617  153.3666   16.22   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
        (Intr) cses
cses    0.000
meanses 0.010  0.000
```

(b) Provide the coefficients of the fixed effects, including their standard errors, and p-values.

Fixed effect	Estimate	SE	p
Intercept γ_{00}	12.68	0.15	< 0.001
Slope of group mean-centred SES γ_{10}	2.19	0.11	< 0.001
Slope of school-avarage SES γ_{01}	5.87	0.36	< 0.001

(c) Specify the same model (from (a)) that tests the contextual effect of school-average SES on studetns' individual mathematics test performance using the R package lavaan.

Provide the model specification and estimate syntax and compute the contextual effect.

Solution:

```
library(lavaan)
# Model specification
mcm <- '
 level: 1
   mAch ~ wa*cses

 level: 2
   mAch ~ wb*meanses

 # Contextual effect
   cont := wb-wa
'
# Model estimation
mcm.fit <- sem(mcm, data=Hsb82, cluster="school")
# Summarise results
summary(mcm.fit, standardized=T, rsq=T)
```

```
lavaan 0.6-5 ended normally after 35 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of free parameters                          5
```

```

  Number of observations                          7185
  Number of clusters [school]                      160

Model Test User Model:

  Test statistic                                 0.000
  Degrees of freedom                                 0

Parameter Estimates:

  Information                                 Observed
  Observed information based on                Hessian
  Standard errors                             Standard


Level 1 [within]:

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  mAch ~
    cses      (wa)    2.191    0.109   20.165    0.000    2.191    0.231

Intercepts:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .mAch              0.000                               0.000    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .mAch             37.014    0.625   59.256    0.000   37.014    0.946

R-Square:
                   Estimate
    mAch              0.054


Level 2 [school]:

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  mAch ~
    meanses   (wb)    5.866    0.359   16.320    0.000    5.866    0.830

Intercepts:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .mAch             12.684    0.148   85.446    0.000   12.684    4.349

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .mAch              2.647    0.397    6.661    0.000    2.647    0.311

R-Square:
                   Estimate
    mAch              0.689

Defined Parameters:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
    cont              3.674    0.375    9.786    0.000    3.674    0.599
```

Contextual effect	Estimate	SE	p
$\gamma_{01} - \gamma_{10}$	3.67	0.38	< 0.001

Cross-level interaction effect of `sector`:

Test the hypothesis that the type of school (`sector`) moderates the relation between students' individual mathematics achievement (`mAch`) and SES (`cses`).

(d) Provide the R code of the model you have used to test this hypothesis in lme4.

Solution:

```
interaction ← lmer(mAch ∼ 1 + cses∗sector + (1 | school), data=Hsb82)
summary(interaction)
```

```
 1 Linear mixed model fit by REML. t-tests use Satterthwaite's method [
 2 lmerModLmerTest]
 3 Formula: mAch ~ 1 + cses * sector + (1 | school)
 4    Data: Hsb82
 5 
 6 REML criterion at convergence: 46650.5
 7 
 8 Scaled residuals:
 9      Min       1Q   Median       3Q      Max 
10 -3.04661 -0.72723  0.01725  0.75624  3.00331 
11 
12 Random effects:
13  Groups   Name        Variance Std.Dev.
14  school   (Intercept)  6.742   2.597  
15  Residual             36.819   6.068  
16 Number of obs: 7185, groups:  school, 160
17 
18 Fixed effects:
19                      Estimate Std. Error        df t value Pr(>|t|)    
20 (Intercept)           11.3930     0.2929  158.4641  38.897  < 2e-16 ***
21 cses                   2.7821     0.1446 7020.0641  19.242  < 2e-16 ***
22 sectorCatholic         2.8053     0.4394  153.7145   6.385 1.93e-09 ***
23 cses:sectorCatholic   -1.3486     0.2184 7020.0641  -6.174 7.02e-10 ***
24 ---
25 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
26 
27 Correlation of Fixed Effects:
28             (Intr) cses   sctrCt
29 cses         0.000              
30 sectorCthlc -0.667  0.000       
31 css:sctrCth  0.000 -0.662  0.000
```

(e) Decide whether this model provides evidence for a cross-level interaction effect and report this effect.

Solution: Line 23 of the R output in (d) reports a t statistic of -6.17 ($p < 0.001$) for the interaction term `cses`$\times$`sector` coefficient, suggesting strong evidence for a cross-level interaction effect.

References

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models for social and behavioural research: Applications and data analysis methods.* Newbury Park, CA: Sage.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic & advanced multilevel modeling* (2nd ed.). London, England: Sage.

Taut, S., Jiménez, D., Puente-Duran, S., Palacios, D., Godoy, M. I., & Manzi, J. (2019). Evaluating the quality of teaching: Can there be valid differentiation in the middle of the performance distribution? *School Effectiveness and School Improvement*, *30*, 328–348. https://doi.org/10.1080/09243453.2018.1510842

Appendix E

Extended Abstract

Overview of Research Topic

Archival data from international large-scale assessments (ILSA) have opened up unique opportunities for examining cross-national trends in equality and equity in educational outcomes. Quasi-longitudinal studies taking advantage of the rich accumulation of ILSA data dating back to the 1960s (Chmielewski, 2019; Majoros et al., 2021) have not only facilitated countries with weak national evaluation or monitoring systems to make comparisons over time, but also enabled comparative studies for all participating states (Salmela-Aro & Chmielewski, 2019; Strietholt & Rosén, 2016) to analyse the impact of specific policy decisions on educational outcomes such as academic de-tracking and expanding education assess. The dual benefits of providing equitable access to policy insight and methodological advancement made systematic studies using historic ILSA data particularly attractive for both policy-makers and educational researchers.

Comparing studies administered decades apart, however, is logistically challenging. Concerns about possible coding errors, for example, is impossible to verify (e.g., FIMS 1964 US sample, Table 1 in Majoros et al. (2021)) due to limitation on early bookkeeping. Secondly, since it is not possible to define a target population balancing both age and grade due to differences in school entry ages across countries, switching target populations from age-based sampling (First International Mathematics Study (FIMS) 1964) to grade-based sampling (Second International Mathematics Study (SIMS) 1980, and subsequent Trends in International Mathematics and Science Study (TIMSS) from 1995 onward) introduced another layer of complexity for comparing different test waves and between ILSAs (Programme for International Student Assessment (PISA) uses age-based sampling). Thirdly, it is questionable whether the measured constructs carried the same meaning over several decades. This problem was particularly acute in early ILSA attempts with some critiquing FIMS for not being a study of mathematics education, but a study of schools and schooling with mathematics serving as a proxy for achievement (Husén (1967), as cited in Majoros et al. (2021)). Lastly, the inherent and stable differences between nation states questioned whether comparisons involving multiple countries is to "compar[e] the incomparable" (Husén (1983), as cited by Kaiser (1999), p.3, then by Majoros et al. (2021), p. 74). All these concerns highlighted the consensus that meaningful comparisons require comparable data.

Key Concepts

SES Achievement Gaps

One fruitful line of enquiry using archival ILSA data linked students' socio-economic status (SES) with their achievement gaps. Using a single-country study design and parental education as the SES measure, Salmela-Aro and Chmielewski (2019) observed a gradual reduction in SES achievement gaps in Finland between 1950s and 1980s but a clear reversal afterwards (Fig 9.2, p. 160). Chmielewski (2019) extended the SES measures using three different variables (a) parental education, (b) parental occupation, and (c) number of books in the household and reported comparable magnitudes of achievement gaps (50, 55 and 40 percent, respectively) across most participating countries. For the purposes of quantifying achievement gaps, existing literature widely adopted the percentile-based approach (Reardon's (2011) method, as cited in Salmela-Aro and Chmielewski (2019)) in which *calibrated* score difference between the 90th and the 10th percentiles of the student cohort were used as an operationalisation for achievement gaps.

Calibration Criteria

In their respective studies, Strietholt and Rosén (2016) and Majoros et al. (2021) both followed Kolen and Brennan (2004, 2014) criteria for evaluating the degree of similarity between tests: inferences, populations, constructs, and measurement characteristics/conditions—referring to low- vs high-stakes tests, age- vs grade-based sampling, terminology shifting, and identical vs complex matrix test designs, respectively.

Causes for Missing Data

When dealing with missing data, Majoros et al. (2021) and Strietholt and Rosén (2016) distinguished three types of missing mechanisms in their studies: not-administrated, omitted, and not-reached items. Not-administrated items were treated as true missing data in estimating item and student parameters while omitted items were treated as incorrect responses in order not to award students for skipping an item. Not-reached items involved more professional judgement with consensus leaning towards recoding them as missing.

Document-type Reading Tasks

Some cycles of reading ILSAs (e.g., RLS-1991/2001) contained unique tasks involving locating information from structured document such as non-continuous tables, chars, graphs, maps or bus timetables. These tasks were later demonstrated to have introduced additional sources of variances into the tests (Gustafsson and Rosén (2006), as cited in Strietholt and Rosén (2016)). Resultantly, all document-type reading tasks were excluded during calibration for the interest of preserving maximum comparability.

Potential Research Questions

Chmielewski (2019) proposed the following research questions:

1. whether increasing SES achievement gaps are a global phenomenon,
2. whether some countries have avoided the trend, and
3. whether increasing SES achievement gaps can be explained by changing educational and social policies and conditions.

Salmela-Aro and Chmielewski (2019) explored the possible drivers behind the initial decline between 1950s and 1980s, following by subsequent increases in SES achievement gaps in Finland. Strietholt and Rosén (2016) and Majoros et al. (2021) both studied in-depth the calibration procedure linking old and recent ILSA data sets.

In addition to mathematics, reading and scientific literacy, interests in "21-Century skills" started to gain momentum in recent decades. Regular surveys of these emerging skills and literacies such as PISA's financial literacy and global competency are also building up their rich libraries of data sources. It is natural to enquire whether the procedural insight gained from multi-decade of traditional literacies applies equally well to 21-Century skills. Potential research question may involve

1. whether SES achievement gaps also appeared in 21-Century skills such as financial literacy,
2. are there any country or country-group (emerging vs industrialised economies) differences in SES achievement gaps in new literacies, and
3. which macro- and microeconomic factors were strongly associated with such systematic variations in SES achievement gaps.

Methodological Approaches

IRT Approach to Calibration

Majoros et al. (2021) and Strietholt and Rosén (2016) applied similar item response theory (IRT) calibration procedures for linking different waves and types of ILSA data sets. Multiple-choice items and dichotomous constructed responses (i.e., 1 mark only with no partial credit given) were subject to 3- and 2-parameter logistic (3PL, 2PL respectively) IRT models to ascertain their difficulty levels. Polytomous items worth 2 or more marks underwent IRT modelling using (generalised) partial credit models (PCM or GPCM) for the same purpose. Standardised marks were then re-scaled to have means 500 and standard deviations 100. Table 3 in Strietholt and Rosén (2016) (p. 11) illustrated the effects of this calibration process by comparing the original and IRT scores.

Other Methodological Considerations

Missing data are the norm rather than the exception in empirical studies, particularly during syntheses of ILSA data sets. Existing publications applied multiple imputation (MI) by iterative chained equations and pulled the five plausible values together following Rubin (1987)'s rules. Recent advancement in both MI theories and software power opened up more options for missing data treatment such as Mplus's unrestricted variance-covariance model using Bayes estimators. The Bayesian procedure may also complement the bootstrap approach to standard error computation employed by existing literature. Mplus's recent upgrade (Version 8.5 and 8.6) combining with hardware infrastructure up to 64-core parallel processing reduced hierarchical growth curve model computation time from days to hours, the multilevel model used in Chmielewski (2019), greatly accelerating incremental model building.

References

Chmielewski, A. K. (2019). The global increase in the socioeconomic achievement gap, 1964 to 2015. *American Sociological Review*, *84*(3), 517–544. https://doi.org/10.1177/0003122419847165

Majoros, E., Rosén, M., Johansson, S., & Gustafsson, J.-E. (2021). Measures of long-term trends in mathematics: Linking large-scale assessments over 50 years. *Educational Assessment, Evaluation and Accountability*, *33*(1), 71–103. https://doi.org/10.1007/s11092-021-09353-z

Salmela-Aro, K., & Chmielewski, A. K. (2019). Socioeconomic inequality and student outcomes in Finnish schools. In L. Volante, S. V. Schnepf, J. Jerrim, & D. A. Klinger (Eds.), *Socioeconomic inequality and student outcomes: Cross-national trends, policies, and practices* (pp. 153–168). Springer. https://doi.org/10.1007/978-981-13-9863-6_9

Strietholt, R., & Rosén, M. (2016). Linking large-scale reading assessments: Measuring international trends over 40 years. *Measurement: Interdisciplinary Research and Perspectives*, *14*(1), 1–26. https://doi.org/10.1080/15366367.2015.1112711

Appendix F

Documentation of English Proficiency: Evidence of an Academic Degree from an English-speaking University

The University of Melbourne

POSTERA CRESCAM LAUDE

This is to certify that

Tony Clifford Austin Tan

was duly admitted to the degree of

Master of Teaching
(Secondary)

in the University of Melbourne on

18 April 2019

Vice Chancellor

Gioconda Di Lorenzo

University Secretary

Appendix G

Documentation of Earlier Education

1. Master of Teaching (Secondary), University of Melbourne, Australia
2. Taxation Law & Practice G (Non-award), University of Canberra, Australia
3. Master of Professional Accounting, Australian National University, Australia
4. Bachelor of Economics, Australian National University, Australia

The University of Melbourne

POSTERA CRESCAM LAUDE

This is to certify that

Tony Clifford Austin Tan

was duly admitted to the degree of

Master of Teaching
(Secondary)

in the University of Melbourne on

18 April 2019

Vice Chancellor

Gioconda Di Lorenzo

University Secretary

ACADEMIC TRANSCRIPT

UNIVERSITY OF CANBERRA
AUSTRALIA'S CAPITAL UNIVERSITY

Mr Tony Clifford Austin Tan

Student ID: 3079861

08 Jun 2012

Non-Award Studies (Examinable)

Course GPA: 7.000

2012

Code	Unit Title	Points	Grade
6279	Taxation Law & Practice G	3.00	HD

End Of Transcript

UNIVERSITY OF CANBERRA
ACT 2601 Australia
www.canberra.edu.au

RESULT GRADES EXPLANATION

HD	High Distinction
DI	Distinction
CR	Credit
P	Pass
UP	Ungraded Pass (pass grade for units assessed on a Pass/Fail basis only)
PX	Conceded Pass (does not meet prerequisite requirements)
NW	Fail result based on written notification of withdrawal from unit after due date
NX	Fail result based on failure to reach pass grade in a unit having completed all the unit assessment requirements
NC	Fail result based on failure to complete one or more of the assessment requirements for the unit
NS	Fail result based on failure to sit for a final examination
NN	Fail result based on non-participation in a unit
WD	Withdrawal approved by Division for good cause
WH	Withheld result - interim result pending finalisation
CNT	Unit continuing over more than one academic year
CNTYL	Unit continuing over one academic year
***	Grade not applicable or not finalised
NAS	Non-Assessment unit
DX	Deferred Examination result pending
SX	Supplementary Examination result pending

HONOURS COURSE LEVELS

H1	First Class Honours
H2A	Second Class Honours, Division I
H2B	Second Class Honours, Division II
H3	Third Class Honours

THE FOLLOWING RESULT GRADES WERE USED PRIOR TO 1998

HD	High Distinction	PM	Pass with Merit (certain units only)
D	Distinction	DX	Deferred Unit result
CR	Credit	NAS	Non-Assessment
P	Pass	KU	Continuing Unit
UP	Ungraded Pass (pass grade for units assessed on a Pass/Fail basis only)	W	Withdrawn
		N	Fail

Note: The University of Canberra was formerly known as the Canberra College of Advanced Education between 1967–1989.

GRADE POINT AVERAGE (GPA)

A GPA is calculated using a seven point scaling system. It is calculated for each course a student undertakes and does not include grades for units completed outside of the course, including units that are credited towards the course. Ungraded Pass (UP) results are not included in a GPA calculation.

ENQUIRIES

Enquiries regarding the contents or the interpretation of this statement should be directed to Student Administration at the address given overleaf or by emailing student.centre@canberra.edu.au

Further information about the University may be found at www.canberra.edu.au

MAJOR AREAS OF STUDY

A major is a sequence of related units in a single discipline or area of study. A major is comprised of between six to eight units. Majors completed by a student will be displayed on the statement at the completion of the course.

SECURITY FEATURES

This Academic Transcript is printed without alterations or erasures, on 100gsm paper.

The micro printing accross the face of the Academic Transcript, which reads "UNIVERSITY OF CANBERRA" under magnification, will not be readable on a copy.

Additionally, a custom made gold foil, featuring the University's C-Star, appears on each page.

PAPER PROFILE

This Academic Transcript is printed on environmentally responsible, 100% recycled paper, independently certified by the Forest Stewardship Council (FSC). No chlorine bleaching occurs in the recycling process. The paper is Certified Carbon Neutral and Greenhouse Friendly by the Australian Government. It is manufactured under the ISO 14001 Environmental Management System.

I certify that the within paperwriting is a true copy of the original document of which it purports to be a copy, such original having been sighted by me. Registrar of the County Court
Dated 11/7/16

KAREN MAVROPOULOS
DEPUTY REGISTRAR OF THE
COUNTY COURT OF VICTORIA
250 WILLIAM ST. MELBOURNE

NATURAM PRIMUM COGNOSCERE RERUM

THE AUSTRALIAN NATIONAL UNIVERSITY

THIS IS TO CERTIFY THAT

FOLLOWING THE COMPLETION OF
AN APPROVED PROGRAM OF STUDY

Yue Tan

HAS BEEN AWARDED THE DEGREE OF

Master of Professional Accounting

GIVEN UNDER THE SEAL OF THE AUSTRALIAN NATIONAL UNIVERSITY
ON THE TENTH DAY OF DECEMBER 2009

Chancellor

Vice - Chancellor

97655

NATURAM PRIMUM COGNOSCERE RERUM

THE AUSTRALIAN NATIONAL UNIVERSITY

THIS IS TO CERTIFY THAT

FOLLOWING THE COMPLETION OF
AN APPROVED PROGRAM OF STUDY

Yue Tan

HAS BEEN AWARDED THE DEGREE OF

Bachelor of Economics

GIVEN UNDER THE SEAL OF THE AUSTRALIAN NATIONAL UNIVERSITY
ON THE SEVENTEENTH DAY OF APRIL 2008

Allan Hawke
Chancellor

Vice-Chancellor

00022961

Appendix H

References

Master Thesis Supervisor

Prof Ronny Scherer
CEMO, University of Oslo
ronny.sherer@cemo.uio.no

Master Thesis Co-supervisor

Dr Chia-Wen Chen
CEMO, University of Oslo
c.w.chen@cemo.uio.no